



HAL
open science

Active listening in sound localization: multisensory and motor contributions to perceiving and re-learning the auditory space

Chiara Valzolgher

► **To cite this version:**

Chiara Valzolgher. Active listening in sound localization: multisensory and motor contributions to perceiving and re-learning the auditory space. Neuroscience. Université de Lyon; Università degli studi di Trento (Trentin, Italie), 2021. English. NNT: 2021LYSE1224 . tel-03662572

HAL Id: tel-03662572

<https://theses.hal.science/tel-03662572>

Submitted on 9 May 2022

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

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



UNIVERSITÀ
DI TRENTO

CiMeC
Center for Mind/Brain Sciences

THESE de DOCTORAT DE L'UNIVERSITE DE LYON

opérée au sein de
l'Université Claude Bernard Lyon 1 et de
University of Trento - Center for Mind/Brain Sciences CIMEC

Écoles Doctorales
ED476 Neurosciences et cognition - NSCo
Doctoral School in Cognitive and Brain Sciences - CIMEC

Spécialité de doctorat : Sciences Cognitives

Discipline : 6 - Département Sciences humaines et humanités - MPSI 01

Soutenue publiquement le 27/10/2021, par :

Chiara Valzolgher

**Active listening in sound localization:
Multisensory and motor contributions to perceiving
and re-learning the auditory space**

Devant le jury composé de :

Isabelle Viaud-Delmon	Institute de Recherche et de Coordination Acoustique/Musique	Rapporteure
Eva Orzan	Ospedale Irccs Materno-Infantile Burlo Garofolo	Rapporteure
Caroline Froment	Centre de Recherche en Neurosciences de Lyon	Examinatrice
Claudio De' Sperati	Università Vita-salute San Raffaele	Examinateur
Alessandro Farnè	Centre de Recherche en Neurosciences de Lyon	Directeur de thèse
Francesco Pavani	Center for Mind/Brain Sciences – CIMEC	Directeur de thèse
Denis Pellisson	Centre de Recherche en Neurosciences de Lyon	Invité
Stefania Benetti	Center for Mind/Brain Sciences – CIMEC	Invitée

This thesis has been prepared at:

ImpAct Team, Centre de Recherche en Neurosciences de Lyon, UMR 5292 - INSERM U1028

16 avenue du Doyen Lépine,

69500 Bron, France

And

Center for Mind/Brain Sciences – CIMeC, University of Trento,

Corso Bettini 13

38068 Rovereto, Italia

Abstract

Sound localization is the ability to perceive the position of auditory events. Humans locate sounds in the environment by interpreting the acoustic signals that reach the ears. When these signals are altered, as in many conditions of hearing loss, locating sounds can become challenging. Adaptation mechanisms enable training of sound localization abilities, even in adulthood. In this thesis, I explored the cognitive mechanisms involved in sound localization and I tested the impact of multisensory and motor variables when training acoustic space perception in normal hearing adults, aging people with presbycusis and cochlear implant users. To these aims, I used an innovative virtual reality approach (VR). In the first two studies, I explored the effect of visual and motor information on sound localization and validated the overall VR approach, in young (Chapter 2) and ageing participants (Chapter 3). In the subsequent three studies, I demonstrated the effectiveness of a reaching-to-sound training, based on multisensory feedback and active listening, in normal hearing people with one ear plugged. First, I showed that this training improves performance more than a comparable control condition (Chapter 4); next, I showed that reaching to sound training generalizes toward a different auditory spatial task in normal hearing people with one ear plugged (Chapter 5); finally, I showed that training and generalization effects can also be observed in people with deafness that use bilateral cochlear implants (Chapter 6). These results extend the current models of acoustic space relearning and propose multisensory-motor trainings that have the potential to transfer to clinical and applied contexts.

Résumé substantiel en français

La localisation du son est la capacité d'identifier l'emplacement des sources sonores dans l'espace 3D. C'est une aptitude extrêmement importante car elle nous permet de diriger notre attention et de percevoir notre environnement. Les êtres humains apprennent au cours de leur vie à localiser les sons sur la base d'indices acoustiques qui dérivent de l'interaction de l'onde sonore avec les oreilles et la tête. En particulier, la localisation des sons nécessite de créer des correspondances entre les signaux auditifs et les coordonnées spatiales de l'espace autour de nous. Notre système cognitif est donc capable de mettre à jour ces correspondances en fonction de la posture et des mouvements de notre corps ainsi que des scénarios acoustiques possibles. Les autres modalités sensorielles, notamment la vue, sont essentielles dans le processus de mise à jour. L'étude de cette capacité de l'être humain est fascinante pour les sciences cognitives car elle permet, d'une part, d'approfondir le caractère multisensoriel de la perception de l'espace acoustique, et d'autre part d'explorer les mécanismes sous-jacents à l'apprentissage de ces correspondances. Il existe en effet des cas où de telles correspondances sont plus difficiles à construire ; les personnes sourdes ayant une récupération auditive partielle, obtenue par exemple grâce à des prothèses auditives ou des implants cochléaires, perçoivent en effet des repères acoustiques souvent altérés voire absents. Pour ces personnes, la localisation des sons devient particulièrement complexe et fatigante, ce qui peut devenir invalidant au quotidien. Il est cependant possible d'induire une amélioration de cette capacité grâce à l'utilisation de rétroactions et d'indices multisensoriels.

Ce travail de thèse traite de la thématique de l'espace acoustique et présente notamment une série d'expériences originales visant à approfondir le rôle de l'écoute active et de l'action directe envers la localisation des sources sonores. Le premier chapitre décrit les mécanismes qui sous-tendent cette capacité et son adaptation aux nombreux changements qui peuvent l'affecter. En particulier, certaines situations dans lesquelles cette capacité est altérée ont été étudiées, comme le cas de l'écoute avec un implant cochléaire. Les chapitres centraux de la thèse consistent en une série de contributions expérimentales – publiées, en cours d'examen ou en préparation – qui ont examiné des questions spécifiques sur le sujet suivant une logique incrémentale. Dans un premier temps, en testant des personnes ayant une audition normale, j'ai validé la méthodologie innovante basée sur la réalité virtuelle, utilisée dans les expérimentations ultérieures de la thèse (Chapitres 2 et 3). Dans un deuxième temps, en utilisant cette même méthodologie, j'ai testé la capacité de localiser les sons des entendants avec une oreille bouchée. Boucher une oreille modifie les signaux acoustiques et constitue donc un modèle d'écoute altérée fréquemment utilisé pour tester l'adaptation dans le temps à de nouveaux signaux acoustiques. Dans ce contexte, j'ai étudié la contribution de la stimulation multisensorielle, du feedback et des actions dirigées vers les sons sur le développement et l'apprentissage de l'espace sonore

correspondant au cours du temps (Chapitre 4). La suite des études précédentes consistait à tester la généralisation des effets d'apprentissage observés (Chapitre 5). Cette généralisation a d'abord été observée chez des entendants ayant une oreille bouchée puis dans un groupe de personnes sourdes porteuses d'implants cochléaires bilatéraux (Chapitre 6). En conclusion, les résultats de ces travaux ont été discutés avec une attention particulière aux apports théoriques qui ont conduit à l'élargissement du modèle de récupération des correspondances entre les indices spatiaux et acoustiques (Chapitre 7).

Cette thèse a démontré le rôle crucial de l'écoute active dans la perception de la position sonore et dans le réapprentissage de cette capacité. Elle a également étendu le modèle théorique de référence en introduisant le concept de stratégie comportementale comme une autre composante à considérer dans l'étude des mécanismes sous-jacents au réapprentissage. Par ailleurs, le rôle critique de l'interaction motrice avec les sons en tant que promoteur du réapprentissage apparaît clairement dans cette thèse. D'un point de vue applicatif, ces recherches offrent une démonstration de l'efficacité d'un entraînement basé sur l'interaction motrice avec le son chez les personnes porteuses d'implants cochléaires bilatéraux. Partant de la validation d'une méthode et de l'étude de modulations spécifiques (visuelles et motrices) sur la localisation des sons, cette série de travaux a conduit à la démonstration de la possibilité d'améliorer la perception acoustique des personnes implantées grâce à un protocole basé sur l'action et l'écoute active mises en œuvre par la réalité virtuelle. Cette thèse présente non seulement un apport théorique mais aussi un outil valable d'intérêt pour le monde clinique, tant pour la mesure que pour la réhabilitation de ces compétences.

Mots clés : Espace acoustique ; Localisation des sons ; Apprentissage ; Mouvements de la tête ; Réalité virtuelle ; Implant cochléaire ; Multisensoriel ; Hypoacousie ; Surdit 

Riassunto sostanziale in italiano

La localizzazione dei suoni è la capacità di identificare la posizione delle fonti sonore nello spazio attorno a noi. Questa abilità è estremamente importante perché permette di percepire l'ambiente circostante nella sua tridimensionalità e influenza i meccanismi di selezione dell'informazione. Gli esseri umani imparano nel corso della vita a localizzare i suoni sulla base degli indizi acustici che derivano dall'interazione dell'onda sonora con le orecchie e la testa. In particolare, localizzare i suoni richiede la creazione di corrispondenze tra i segnali uditivi che raggiungono le orecchie e le coordinate tridimensionali dello spazio esterno. Il nostro sistema cognitivo è in grado di aggiornare queste corrispondenze in funzione della postura e dei movimenti del corpo, adattandole agli scenari acustici di volta in volta disponibili. Le altre modalità sensoriali, in particolare la vista, sono essenziali nel processo di aggiornamento. Studiare questa abilità dell'essere umano è affascinante per le scienze cognitive perché permette di approfondire la natura multisensoriale della percezione dello spazio acustico e consente di indagare i meccanismi sottostanti all'apprendimento di tali corrispondenze. Esistono infatti dei casi in cui tali corrispondenze sono più difficili da costruire. Ad esempio, le persone sorde con un parziale recupero uditivo, ottenuto grazie a protesi acustiche e impianti cocleari, percepiscono indizi acustici spesso alterati. Per queste persone, localizzare i suoni diventa particolarmente complesso e faticoso, e le conseguenze di questa difficoltà possono diventare invalidanti nella vita di tutti i giorni.

Il presente lavoro di tesi tratta il tema dello spazio acustico e in particolare porta una serie di contributi sperimentali originali volti ad approfondire il ruolo dell'ascolto attivo e dell'azione diretta verso i suoni nella loro localizzazione. Nel primo capitolo sono descritti i meccanismi che sottendono questa abilità e come possano adattarsi ai numerosi cambiamenti che possono influenzarla. In particolare, sono state approfondite alcune situazioni in cui questa abilità viene alterata, come il caso di ascolto con impianto cocleare. I capitoli centrali della tesi sono costituiti da una serie di contributi sperimentali – pubblicati, inviati o in preparazione – che hanno affrontato domande sperimentali specifiche seguendo una logica incrementale. Dapprima, è stata validata la metodologia innovativa basata sulla realtà virtuale impiegata in tutti gli esperimenti della tesi testando persone con udito normale (Capitolo 2 e 3). Successivamente, utilizzando tale metodologia, sono state testate le abilità di localizzare i suoni di ascoltatori con un orecchio tappato. Tappare un orecchio modifica gli indizi acustici e pertanto costituisce un modello di ascolto alterato frequentemente utilizzato per testare sperimentalmente la possibilità di adattamento a dei nuovi indizi acustici nel corso del tempo. In questo contesto, è stato indagato il contributo della stimolazione multisensoriale, del feedback e in particolare delle azioni dirette verso i suoni sullo sviluppo e l'apprendimento delle corrispondenze suono-spazio nel corso del tempo (Capitolo 4). Un ulteriore passo in avanti rispetto agli studi precedenti, è stato testare la generalizzazione degli effetti di apprendimento

osservati (Capitolo 5). La generalizzazione è stata prima osservata nelle persone udenti con un orecchio tappato e poi in un gruppo di persone sorde con impianto cocleare bilaterale (Capitolo 6). A conclusione, sono stati discussi i risultati dei lavori sperimentali con particolare riferimento ai contributi teorici che hanno portato ad espandere il modello di riapprendimento delle corrispondenze tra spazio e indizi acustici già presente in letteratura (Capitolo 7).

Questo lavoro di tesi ha dimostrato il ruolo cruciale dell'ascolto attivo nella percezione della posizione del suono e nel riapprendimento di questa capacità. Ha inoltre esteso il modello teorico di riferimento introducendo anche il concetto di strategia comportamentale come ulteriore componente da considerare nello studio dei meccanismi sottostanti il riapprendimento. Nel presente lavoro di tesi emerge chiaramente il ruolo cruciale dell'interazione motoria con i suoni come promotore del riapprendimento. Da un punto di vista applicativo, queste ricerche offrono una dimostrazione dell'efficacia di un addestramento basato sull'interazione motoria con il suono nelle persone portatrici di impianto cocleare bilaterale. Partendo dalla validazione di un metodo e dallo studio di specifiche modulazioni (visive e motorie) sulla localizzazione dei suoni, questa serie di lavori ha portato a dimostrare la possibilità di migliorare la percezione acustica delle persone con impianto cocleare attraverso un protocollo basato sull'azione e l'ascolto attivo implementato in realtà virtuale, portando quindi non solo un contributo teorico, ma anche un valido strumento di interesse per il mondo clinico sia per la misurazione che per la riabilitazione di tali abilità.

Parole chiave: Spazio acustico; Localizzazione dei suoni; Apprendimento; Movimenti della testa; Realtà virtuale; Impianto cocleare; Multisensorialità; Ipoacusia; Sordità

Summary in English

Sound localization is the ability to identify the location of sounds in space. This skill is extremely important because it allows to perceive the three dimensions of the auditory environment and contributes to attentional orienting and selection. Human learn over the course of life to localize sounds by relying on auditory cues that derive from the interactions of sound waves with the ears and head. Localizing sounds requires creating correspondences between auditory signals and spatial coordinates. Our cognitive system can update these correspondences according to posture and body movements as well as the multiple acoustic scenarios we encounter in life. The other sensory modalities, especially vision, are essential in this updating process. Studying this ability is fascinating for cognitive sciences because it allows to examine the multisensory nature of the acoustic space perception and it permits the investigation of the learning mechanisms subtending these spatial skills. For some people such correspondences are more difficult to create. For instance, people with deafness who use hearing aids or cochlear implants must reconstruct these correspondences through altered and partial auditory cues. For these people, localizing sounds becomes particularly complex and effortful, with disabling consequences in everyday life.

This thesis deals with acoustic space perception and presents a series of original experimental contributions aimed at examining the role of active listening and sound-directed actions in spatial hearing. The first chapter describes the mechanisms underlying sound localization and how they can adapt to altered listening conditions, such as the case of listening with one ear plugged or through cochlear implants. The central chapters of the thesis present a series of experimental contributions – published, submitted, or in preparation - which investigated specific questions about acoustic space perception following an incremental logic. First, I validated the innovative methodology based on virtual reality used in all subsequent experiments of the thesis in people with normal hearing (Chapters 2 and 3). Using this methodology, I tested the ability to locate the sounds in listeners with one ear plugged. Plugging one ear modifies the auditory cues and therefore it constitutes a frequently used model to test adaptation to new auditory cues over time. Second, using this model, I investigated the contribution of multisensory stimulation, feedback and sound-directed actions to the adaptation to altered auditory cues over time (Chapter 4). A further step forward from previous studies was to test the generalization of observed learning effects (Chapter 5). The generalization was first observed in normal hearing with one ear plugged and then in a group people with deafness using bilateral cochlear implants (Chapter 6). In the concluding chapter, the theoretical and applied contributions of the experimental works were discussed. They expand the model of acoustic space relearning already present in the literature (Chapter 7).

This thesis demonstrates the crucial role of active listening in acoustic space perception and relearning. The concept of behavioral strategy is presented as a further component to consider when studying the mechanisms underlying relearning. Furthermore, the crucial role of motor interaction with sounds as a promoter of relearning clearly emerges in this thesis. From an applied perspective, the series of studies presented in the thesis offer a demonstration of the effectiveness of training based on motor interactions with sound in people with bilateral cochlear implants. Starting from the validation of new methodological approach and by testing specific modulations (visual and motor) on the localization of sounds, this thesis dissertation led to the demonstration of the possibility of improving the acoustic perception of people with cochlear implants through a protocol based on sound-oriented actions and active listening, implemented in virtual reality. In this respect, this offers theoretical contributions as well as practical indications for building a training approach that could transfer to the clinical world both for the measurement and for the rehabilitation of auditory spatial skills.

Key words: Acoustic space; Sounds localization; Learning; Head-movements; Virtual reality; Cochlear implant; Multisensory; Hard-of-hearing; Deafness

Table of contents

Preview.....	20
Chapter 1. Background.....	23
1.1. Perceiving the auditory space	23
1.1.1. Auditory cues to localize sounds	23
1.1.2. From auditory cues to spatial representations.....	27
1.1.3. Sound localization in a visual world.....	31
1.1.4. Head movements and spatial hearing.....	34
1.2. Re-learning sound localization	36
1.3. Toward a clinical model.....	41
Chapter 2. Study 1: The impact of a visual spatial frame on real sound-source localization in virtual reality	46
2.1. Abstract.....	47
2.2. Introduction	47
2.3. Methods	50
2.3.1. Participants	50
2.3.2. Apparatus and stimuli.....	50
2.3.3. Procedure	51
2.3.4. Analyses	54
2.4. Results.....	54
2.4.1. Hand pointing responses: the uniform grey condition	55
2.4.2. Hand pointing responses: the effects of adding a visual frame	58
2.4.3. Head rotation.....	60
2.4.4. Gaze direction.....	63
2.5. Discussion.....	63
2.6. Conclusion	67

Chapter 3.	Study 2: Spontaneous head-movements improve sound localization in ageing adults with presbycusis	68
3.1.	Abstract.....	69
3.2.	Introduction	69
3.3.	Material and Methods.....	71
3.3.1.	Participants	71
3.3.2.	Stimuli	72
3.3.3.	Apparatus.....	72
3.3.4.	Procedure	73
3.3.5.	Analysis.....	76
3.4.	Results.....	77
3.4.1.	Front-back discrimination task.....	77
3.4.2.	Sound localization task in the frontal plane	78
3.5.	Discussion.....	79
Chapter 4.	Study 3: Reaching to sounds in virtual reality: A multisensory-motor approach to promote adaptation to altered auditory cues	82
4.1.	Abstract.....	83
4.2.	Introduction	83
4.3.	Material and Methods.....	88
4.3.1.	Participants	88
4.3.2.	Apparatus and stimuli.....	88
4.3.3.	Procedure	90
4.3.4.	Head-movement pre-processing	91
4.4.	Results.....	91
4.4.1.	Sound localization performance	91
4.4.2.	Head movements	95
4.4.3.	Relation between changes in sound localization performance and head-movements ...	98
4.4.4.	VR experience.....	99
4.4.5.	Sense of agency	100

4.4.6.	Gaming experience	100
4.5.	Discussion.....	101
4.5.1.	Sound localization improves rapidly during monaural listening	102
4.5.2.	Monaural listening triggered compensatory head-movements	103
4.5.3.	Reaching to sounds promote adaptation to monaural listening more than naming	105
4.6.	Conclusions	107
Chapter 5.	Study 4: Adapting to altered auditory cues: generalization from manual reaching to head pointing	108
5.1.	Abstract.....	109
5.2.	Introduction	109
5.3.	Methods	111
5.3.1.	Participants	111
5.3.2.	General structure of the experimental session	112
5.3.3.	Apparatus.....	114
5.3.4.	Procedure and stimuli	114
5.3.5.	2.4.1. Testing Phase	115
5.3.6.	Training Phase	116
5.3.7.	Analysis.....	118
5.4.	Results.....	119
5.4.1.	Does monaural listening increase sound localization errors?.....	119
5.4.2.	Does spatial training reduce sound localization errors?	121
5.4.3.	Does Spatial training effects generalize to the head-pointing sound localization task?	122
5.4.4.	Do improvements during the spatial training predict generalization effects?	125
5.4.5.	Does Spatial training change spontaneous head-rotation behavior while listening?	126
5.4.6.	Does training effects change immediate head-orienting to sounds?	127
5.4.7.	Does training induce changes in audio-visual orienting?	129
5.5.	Discussion.....	129
5.5.1.	Reaching to sound is an effective training strategy	130

5.5.2.	A possible role for head-movements in adapting to altered spatial hearing	132
5.6.	Conclusions	133
Chapter 6.	Study 5 - Reaching to sounds improves spatial hearing in bilateral cochlear implant users. A Randomized Clinical Trial.	135
6.1.	Abstract.....	136
6.2.	Introduction	136
6.3.	Materials and methods.....	138
6.3.1.	Study Design.....	138
6.3.2.	Participants	140
6.3.3.	Procedures.....	141
6.3.4.	Statistical Analysis	142
6.4.	Results.....	142
6.4.1.	Changes in sound localization and head-movements during Spatial training.....	142
6.4.2.	Effects of Spatial training on head-pointing sound localization.....	144
6.4.3.	Audio-Visual Cueing task	146
6.5.	Discussion.....	146
6.6.	Conclusion	147
6.7.	Supplementary materials	148
6.7.1.	Supplementary materials and methods.....	148
6.7.2.	Supplementary Results	154
Chapter 7.	General Discussion	157
7.1.	Answers to the experimental questions.....	157
7.1.1.	Seeing an empty visual environment improves sound localization in elevation and produces faster gaze orienting responses to the sound.....	157
7.1.2.	Head movements improve sound localization, also in ageing people with presbycusis	158
7.1.3.	Reaching to sounds helps adaptation to altered auditory cues more than naming the sources	158
7.1.4.	The effects of sound localization by reaching generalize to a different sound localization task	159

7.1.5.	Sound localization in bilateral cochlear implant users can be trained though reaching to sounds	160
7.2.	Further elements for the conceptual model of acoustic space relearning	160
7.2.1.	Expanding the model: the contribution of active listening.....	163
7.3.	Experimental work on listening require a paradigm shift.....	165
7.3.1.	Measuring spatial hearing should occur in the three dimensions of space.....	165
7.3.2.	Active listening should be allowed and measured.....	167
7.3.3.	Control of the visual scenarios and multisensory cues to sound position is essential ..	168
7.3.4.	A paradigm shift for experimental and clinical studies.....	168
7.4.	Limitations and future directions.....	169
7.4.1.	Clarify the role of behavioral strategies	169
7.4.2.	Including listening effort measures.....	170
7.4.3.	Expanding towards metacognitive belief	171
7.4.4.	Triggering efficient behavioral strategies.....	172
7.4.5.	Thinking about internal feedback and confidence	172
7.4.6.	Enriching the acoustic space perception model	173

Figures list

FIGURE 1 AUDITORY CUES.....	25
FIGURE 2 SCHEMATIC OF HUMAN AUDITORY SYSTEM.....	27
FIGURE 3 CORTICAL NETWORK OF ACTIVE, GOAL-ORIENTED SOUND LOCALIZATION PROCESSING	30
FIGURE 4 CORRELATION BETWEEN SOUND LOCALIZATION THRESHOLD AND WIDTH OF FIELD OF BEST VISION ...	31
FIGURE 5 SCHEMATIC DESCRIPTION OF A COCHLEAR IMPLANT.....	44
FIGURE 6 STUDY 1 - EXPERIMENTAL PROCEDURE AND SETTING	51
FIGURE 7 STUDY 1 - SOUND LOCALIZATION PERFORMANCE.....	56
FIGURE 8 STUDY 1 - SIGNED ERROR IN ELEVATION.....	59
FIGURE 9 STUDY 1 - HEAD ROTATION	61
FIGURE 10 STUDY 1 - HEAD AND GAZE FIRST MOVEMENTS	62
FIGURE 11 STUDY 2 - FRONT-BACK DISCRIMINATION SETTING.....	75
FIGURE 12 STUDY 2 - SOUND LOCALIZATION SETTING	76
FIGURE 13 STUDY 2 - FRONT-BACK DISCRIMINATION PERFORMANCE.....	78
FIGURE 14 STUDY 2 - SOUND LOCALIZATION PERFORMANCE.....	79
FIGURE 15 STUDY 3 - EXPERIMENTAL PROCEDURE AND SETTING.....	87
FIGURE 16 STUDY 3 - SOUND LOCALIZATION PERFORMANCE.....	92
FIGURE 17 STUDY 3 - SOUND LOCALIZATION PERFORMANCE AS FUNCTION OF TRIAL	94
FIGURE 18 STUDY 3 - SLOPE AND INTERCEPT INDECES CALCULATION	95
FIGURE 19 STUDY 3 - NUMBER OF MOVEMENT AND AMPLITUDE	96
FIGURE 20 STUDY 3 - SPACE EXPLORED	97
FIGURE 21 STUDY 3 - CORRELATION BETWEEN SPACE EXPLORED ROTATING THE HEAD AND PERFORMANCE	99
FIGURE 22 STUDY 3 - VR EXPERIENCE.....	101
FIGURE 23 STUDY 4 - EXPERIMENTAL PROCEDURE AND SETTING	113
FIGURE 24 STUDY 4 - HEAD-POINTING SOUND LOCALIZATION IN SESSION 1 BEFORE TRAINING.....	120
FIGURE 25 STUDY 4 - HEAD-POINTING SOUND LOCALIZATION, SESSION 1	121
FIGURE 26 STUDY 4 - PERFORMANCE DURING SPATIAL AND NON-SPATIAL TRAINING.....	122
FIGURE 27 STUDY 4 - PERFORMANCE IN HEAD-POINTING SOUND LOCALIZATION	123
FIGURE 28 STUDY 4 - CORRELATION	126
FIGURE 29 STUDY 4 - CHANGES IN HEAD BEHAVIOUR DURING THE SPATIAL TRAINING	127
FIGURE 30 STUDY 4 - FIRST HEAD-MOVEMENTS REACTION TIME.....	128
FIGURE 31 STUDY 4 - AUDIO-VISUAL CUEING TASK.....	129
FIGURE 32 STUDY 5 - EXPERIMENTAL PROCEDURE AND SETTING.....	139
FIGURE 33 STUDY 5 - PARTICIPANTS' PREFORMANCE DURING TRAINING	143
FIGURE 34 STUDY 5 - PERFORMANCE IN HEAD-POINTING SOUND LOCALIZATION	145
FIGURE 35 STUDY 5 - HEAD MOVEMENT IN THE SPATIAL TRAINING	155
FIGURE 36 STUDY 5 - HEAD MOVEMENT IN IN HEAD-POINTING SOUND LOCALIZATION	156
FIGURE 37 MODEL OF AUDITORY ADAPTATION PROCESSING.....	163

Tables list

TABLE 1 SCHEMATIC REPRESENTATION OF EXPERIMENTAL QUESTIONS AND RELATED STUDIES	22
TABLE 2 STUDY 1 - ABSOLUTE ERROR AND SIGNED ERROR	60
TABLE 3 STUDY 2 - ANAMNESTIC AND AUDITORY CHARACTERISTICS OF PARTICIPANTS	72
TABLE 4 STUDY 2 - ACTIVE AND STATIC ERROR	79
TABLE 5 STUDY 4 - PERFORMANCE IN HEAD-POINTING SOUND LOCALIZATION	124
TABLE 6 STUDY 4 - RESULTS OF THE LME ANALYSES OF AZIMUTH AND ELEVATION ABSOLUTE AND SIGNED ERRORS	125
TABLE 7 STUDY 5 - BIOGRAPHIC AND ANAMNESTIC AND CLINICAL CHARACTERISTICS OF CI PARTICIPANTS	140
TABLE 8 STUDY 5 - EXTENDED INFORMATION ABOUT PARTICIPANTS' COCHLEAR IMPLANTS	149

Preview

Knowing where things are in space is an ordinary multisensory experience. When we see our phone, we can clearly identify its spatial position through vision. We see that it is placed on the table, a meter from the ground, on our right, about 50 centimeters away from us – that is, we encode its position in a three-dimensional (3D) reference frame. Furthermore, when our phone is hidden from view, we can use its sound to track its location. Finally, we can plan a motor action to reach it and feel its vibration on the hand's skin. Information about where things are in the space is fundamental to perceive the 3D multisensory environment and interact with objects.

Sounds localization is the ability to identify sound sources in 3D space and it is extremely important to our everyday life. It is not only tied to the ability to identify the location of a single source, but it also entails the ability of locating this source with respect to ones' own body (i.e., an egocentric reference frame encoding) and to capture its spatial relationship with respect to other sound sources or events (i.e., an allocentric reference frame encoding). Critically, sound localization requires learning correspondences between auditory cues (at one or both ears) and spatial coordinates of the external space. Unlike visual and somatosensory systems (Middlebrooks & Green, 1991), the auditory system needs to interpret auditory cues resulting from the interaction between sound waves, the head, and the external ears when inferring precise coordinates in space. Therefore, knowing where sounds come from (and consequently where sound sources are in space) is the result of a complex process and it represents one of the extraordinary abilities of our brain. Before delving into the mechanisms and the yet to be discovered issues of acoustic space, I will spend a few words to describe why localizing sounds is so important.

The first reason is that spatial hearing plays a key role when monitoring the surrounding environment. It is possible to perceive acoustic information coming from far away or from behind, well beyond the available visual space. Two aspects make spatial hearing particularly relevant to this aim: first, audition is distal sense and thus it allows to perceive a stimulus before it reaches the body; second, sounds are only partially limited by obstacles and can travel long distances in space. Understanding the origin of an animal sound or noise is extremely useful for hunting or, conversely, for localizing potential predators. The ability to figure out from where its predator comes, gives an animal the opportunity to implement efficient escape or avoidance reactions (Heffner & Heffner, 2014; Nothwang, 2016).

The second reason is closely related to the previous but entails the other sensory systems. Once an event is detected through hearing, the organism can seek for more information about it using the other available senses. With respect to spatial hearing, in particular, it favors orienting of multisensory attention in space,

allowing re-orienting of the receptors and encoding through other modalities (Heffner & Heffner, 1992; Pavani et al., 2017).

Sounds localization plays a crucial role in the interpretation of the auditory scene and in the formation of the auditory objects (Middlebrooks, Simon, Popper, & Fay, 2017; Bregman, 1994) and it helps to discern the signal from noise (Yost, 2017). The third reason refers to its link with human communication. Sound localization is an advantage when people communicate. Imagine the typical situation of a cocktail party (Cherry, 1953). Localization of sound sources means identifying each acoustic stream in space and facilitating the segregation of the auditory objects from the acoustic mixture that has reached our ears. To extract the voice of a determined speaker from mixed auditory traces, the brain must identify it, separating it from other auditory objects in the scene and allocating attentional resources to it (Shinn-Cunningham, Best, & Lee, 2017). This task relies both on the quality of the signal reaching the ear and then transmitted to the brain, and on the efficiency of cognitive mechanisms such as auditory processing, working memory, selective attention, and precisely spatial perception.

Here, I will examine spatial perception of sounds considering it as a multisensory and plastic ability of our cognitive system. The thesis is organized in 7 chapters. In Chapter 1, I will describe the mechanisms that subtend perception of acoustic space, together with some of the open experimental questions in this research domain. These queries will be investigated and discussed further through a series of empirical chapters (Chapter 2 to 6) which will present the studies I have conducted during the PhD. Finally, in Chapter 7 I will summarize the answers that originated from the experimental work, the contributions that they led to the overall theoretical framework, and the future directions that emerged while working on this issue.

The background chapter (Chapter 1) is divided into three sections. Section 1.1 will focus on the mechanisms subtending human ability to localize sounds. In particular, I will describe how the auditory system can extract auditory spatial representations starting from the initial sound waves reaching the ears. Furthermore, the contribution of visual information when localizing sound will be discussed, together with the role of head movements. Section 1.2. will be dedicated to our brain's ability to adapt to different listening conditions. Sound localization is an essential skill for humans as well as other hearing species, to the point that it can adapt to the many changes that can influence its functionality. These include changes in the body of the listener (e.g. growth in the size of the head or ears), changes in the environment (e.g. room reverberation, out-door spaces) or alterations in the availability of the auditory cues reaching the ears (e.g. forms of hearing loss). In this section, I will give a theoretical background about how our brain can adapt to this altered information to continue to perceive acoustic space efficiently. The most relevant examples of alteration of auditory cues are constituted by hearing problems that can occur over the course of life, such as presbycusis. Moreover, the widely used technological devices for hearing

recovery also alter these auditory cues. This is the case, for example, of cochlear implants. In section 1.3, I will focus on more practical and clinically relevant examples of acoustic space perception alteration.

The core part of my work will present articles published, submitted, or in preparation on which I have worked on during my PhD. A schematic representation of the studies and the main experimental questions is presented below (Table 1). All the experimental contributions presented shared the same methodological approach. I took advantage of a virtual reality innovative system that combines real sounds presentation and motion tracking. The peculiarities and the advantage of this method will be discussed at session 7.3.

Title	Chapter	Main experimental question	Population
The impact of a visual spatial frame on real sound-source localization in virtual reality	2	Does seeing an empty visual environment impact on spatial hearing?	Normal Hearing – Binaural listening
Spontaneous head-movements improve sound localization in ageing adults with presbycusis	3	Do head movements improve sound localization in ageing people with presbycusis?	Asymmetrical Presbycusis
Reaching to sounds in virtual reality: A multisensory-motor approach to promote adaptation to altered auditory cues	4	Does sound localization by reaching promote adaptation to altered auditory cues more than just naming the sources?	Normal Hearing – binaural and monaural listening (one-ear plugged)
Adapting to altered auditory cues: generalization from manual reaching to head pointing	5	Does sound localization by reaching train explicit and implicit sound localization abilities?	Normal Hearing – binaural and monaural listening (one-ear plugged)
Reaching to sounds improves spatial hearing in bilateral cochlear implant users A Randomized Clinical Trial	6	Can we train sound localization in bilateral cochlear implant users?	Bilateral Cochlear implant users

Table 1 SCHEMATIC REPRESENTATION OF EXPERIMENTAL QUESTIONS AND RELATED STUDIES

The general discussion (Chapter 7) is divided into 4 sections. Sections 7.1 sums up the series of studies proposed, discussing more broadly the aspects which can contribute to enrich the current state of the art in the domain of acoustic space perception and training. The 7.2 deepens contributions of the thesis's finding to the model of acoustic space relearning. The 7.3. discusses the advantages of the methodological approach adopted across experimental contributions. Finally, the 7.4 is dedicated to limits and future directions.

Chapter 1. Background

1.1. Perceiving the auditory space

1.1.1. Auditory cues to localize sounds

Sounds around us are waves of energy travelling from the source that generated them all the way to our ears. When they enter the ear canal they make the eardrum vibrate and trigger the chain of events that eventually results in perceived sounds. Sound waves hit the eardrum membrane, causing a one-dimensional movement, irrespective of the origin or complexity of the sound (Grothe, Pecka, & McAlpine, 2010). This vibration travels in the form of mechanical energy all the way to the cochlea, through the three auditory ossicles – the malleus, the incus and the stapes – in the middle ear. The auditory ossicles transfer and amplify the vibration received at the eardrum, exerting pressure on the oval window, a membrane on the cochlea. The amplified vibrations (the pressure gain is of about 20 times the sound pressure in the air) are thus transferred to the fluids which fill the cochlea. This liquid displacement is reflected in stress on the basilar membrane, which is attached to the base of the cochlea. On this membrane, there is the organ of Corti, which contains the sensory receptors of the auditory system: the hair cells. Hair cells are clustered into outer cells, which amplify mechanically the signal and are involved in sound sensitivity and frequency discrimination, and inner cells, responsible for the actual transformation of sound vibration in the fluids into electrical signal (i.e., the transduction process). On their top, inner cells present special structures, called stereocilia, which move following the endolymph movements. Through mechanotransduction, stereocilia transform the mechanic energy into electrical signals. Thus, hair cells transmit this signal to neurons, the auditory nerve, which convey the information to the brain (Bear, Connors, & Paradiso, 2020).

The spatial distribution of hair cells in the cochlea follows a tonotopic organization, meaning that they are organized according to sound frequencies (20-20000 Hz in humans). Cells at the base of the cochlea respond to higher-pitched sounds, whereas the ones toward the top of the cochlea respond to lower-pitched sounds. The tonotopic organization is a fundamental characteristic of the auditory system, but does not convey information about the spatial position of the sound. The hair cells are positioned along the basilar membrane according to the frequency of the sounds rather than the sounds' origin (Humphries, Liebenthal, & Binder, 2010; Saenz & Langers, 2014). There is no direct correspondence on the cochlear surface between the space of the sound in the external world and the topographical distribution of the receptors. This is markedly different with respect to the visual and somatosensory systems, which present a retinotopic and somatotopic organization respectively (Bear, Connors, & Paradiso, 2020) and can thus map some of the spatial features of visual and somatosensory events already at the receptor's surface.

Given this peculiarity of the auditory system, the cognitive system needs to interpret spatial auditory information without any explicit representation of auditory space on the sensory receptors and also on the primary sensory cortices (Grothe et al., 2010).

Spatial hearing requires the creation of an acoustic map through the learning of correspondences between auditory cues (provided by both ears) and spatial coordinates of the external space. To do this, the auditory system needs to interpret the auditory cues resulting from the interaction between sound waves, the head and the external ears (Middlebrooks, 2015). While some aspects of this auditory cues-space association are present since birth (Muir, Clifton & Clarkson, 1989; Németh, Háden, Török, & Winkler, 2015), this ability matures over a longer period of time (Litovsky, Ashmead, Gilkey, & Anderson, 1997). During life, people learn to interpret these cues, which vary as a function of both sound sources position and one's own body posture, and to associate them to precise spatial positions, which can be described considering the three dimensions of the space: azimuth (horizontal plane), elevation (vertical plane), and distance. The plasticity of the cognitive system is fundamental for spatial hearing because it permits to adapt and relearn to interpret auditory cues taking into account the external environment, the dynamic changes in one's own body and head posture (Chechik, Meilijson, & Ruppin, 1999; Clifton, Gwiazda, Bauer, Clarkson, & Held, 1988) and the cues provided by other modalities such as the visual cues (Keating & King, 2015).

The auditory cues that the cognitive system can extract from the sounds reaching the ears are critical to obtain spatial information. Each cue can provide information about one or more spatial dimensions of the source. We distinguish between binaural cues, which result from the combination of auditory inputs in both ears, and monaural cues, which refer to the information available at every single ear, such as the specific spectral composition of the sound which reaches the single ear. While binaural cues allow localization in the horizontal dimension, monaural cues are more crucial to identify sound position in elevation and they contribute to front-back disambiguation and distance estimation (Middlebrooks, 2015; Middlebrooks & Green, 1991; Risoud et al., 2018) (Figure 1).

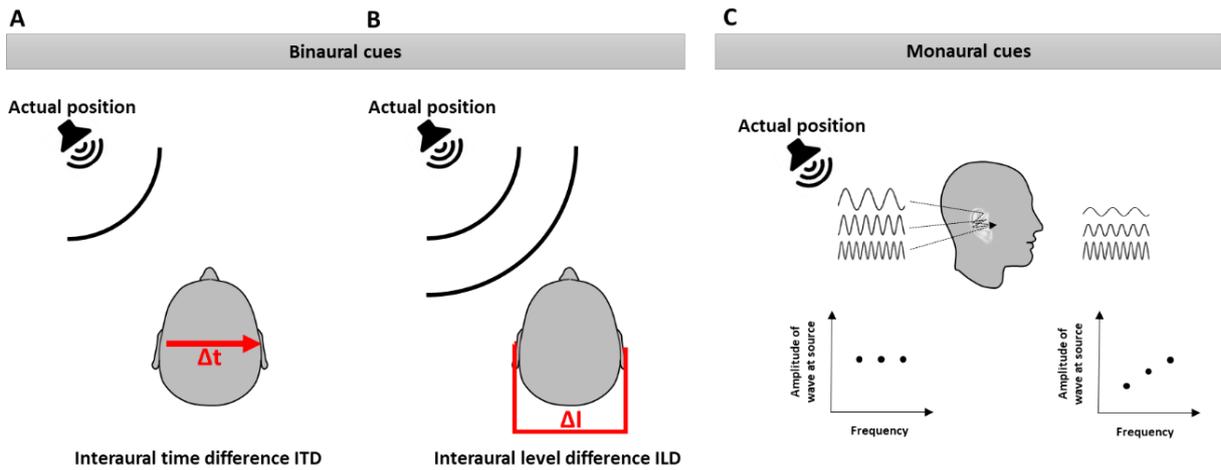


Figure 1 AUDITORY CUES

Schematic representations of binaural cues: Interaural time difference ITD (A) and Interaural level difference ILD (B) and monaural cues (C). The drawn was inspired and adapted from Groh, 2014 and Grothe et al., 2010.

Binaural cues are distinguished in interaural time differences (ITDs) and interaural level differences (ILDs). ITDs result from the fact that sound waves from a certain source reach the closer ear before the more distant one. The temporal difference in the arrival of the signal at the two ears is detectable from our cognitive system (Brughera, Dunai, & Hartmann, 2013) and it is particularly useful for localization of low-frequency sounds (< 1400 Hz), whose wavelengths are larger than the diameter of the head. ILDs are determined by the shadowing effect of the head: sounds are louder in the ear nearer to the source compared to the other one. Higher frequency sound waves, whose wavelengths are shorter than the head's dimension (> 4000 Hz), when colliding with the head determine the acoustic shadow because the head acts as an obstacle. Consequently, it creates a great difference in sound loudness between the two ears. Humans can detect up to 0.5 dB disparity (Van Opstal 2016). Both ILDs and ITDs are informative when we localize sound in the horizontal plane, however, their efficacy depends on the frequency of the sound waves (Macpherson & Middlebrooks, 2002). Differently from the ITDs, ILDs are affected by source distance: when sounds come further away 1 or 2 meters from the listeners they are imperceptible, while for nearby sounds they increase for lateral source significantly (Brungart & Rabinowitz, 1999; Shinn-Cunningham, Santarelli, & Kopco, 2000). This happens because our brain takes advantage of the difference between the lengths of the paths from the source to each ear. This difference decreases when the distance of the source from the head increases and it is also ambiguous when the source is located on the medial plane (i.e. same distance between the source and both ears) (Shinn-Cunningham, Santarelli & Kopco, 2000).

Monaural cues are the information available at every single ear. They are mainly involved in the identification of the spatial position of sounds in the vertical plane, in depth, and in the anterior-posterior axis. When sound waves reach the ear, their acoustic spectrum is modified by the contact with the body,

the head and the external ear (i.e. the conformation of the pinna). These changes in the spectral configuration occur as a function of the direction of the sound reaching the eardrum. They are unique for each person because of the various features of each body, and ears (shape, size). These alterations in the sound spectrum are termed Head Related Transfer Functions (HRTFs) (Carlile, Martin, & McAnally, 2005; Colburn & Kulkarni, 2005). Monaural cues are extracted separately for both ears and then integrated. Interestingly, this integration could be partially asymmetrical, as in judging the elevation of a certain sound source, a major weight is given to the ear ipsilateral to the source (Van Wanrooij & Van Opstal, 2005).

Monaural cues are also involved in the estimation of distance: higher frequencies content decreases as a function of distance. Stimuli in which lower frequencies predominate over higher ones are perceived as more distant (Musicant & Butler, 1980). However, monaural cues are not enough to localize sounds' distance. Our cognitive system needs to combine them with ILDs, which vary especially when sources are near (< 1 meter), ITDs, and the direct-to-reverberant energy ratio (DRR). The latter parameter refers to the relationship between the amount of sound energy that reaches our ears directly from the source and the amount which was previously reflected on surrounding surfaces (Ahveninen, Kopčo, & Jääskeläinen, 2014; Kolarik, Moore, Zahorik, Cirstea, & Pardhan, 2016).

As with distance estimation, there are other ambiguous situations in which binaural and monaural cues must be integrated to extract the correct position of the sound. An example in which this integration is necessary is the case of the "cone of confusion", a cone-shaped area extending from the ear (Shinn-Cunningham, Santarelli, & Kopco, 2000). In this specific area, ILDs and ITDs are the same for all the positions. Thus, individuals cannot identify whether the sound is coming from the front or the backspace. Spectral cues are necessary to disambiguate between sounds at the front and sounds at the back. The posterior part of the pinnae filters the sound when it comes from the back, while the more frontal part of the ear changes the sound when it comes from the front.

These examples of cues integration highlight a key notion that will become particularly relevant in the following sections of the thesis: combining binaural and monaural information is important even in typical human listeners, and the outcome of process influences the accuracy of sound localization. Most importantly, this plastic ability to integrate information is functional to adapt to the dynamicity of the environment and to hearing alteration (see 1.2.) (e.g. Keating, Dahmen, & King, 2013).

In sum, spatial hearing builds from the auditory cues available at the ears. However, to fully understand how these cues are integrated with spatial representation in the brain it is important to follow the auditory pathway from the ear to the auditory cortex and beyond.

1.1.2. From auditory cues to spatial representations

Acoustic information from the ear reaches the cortex via the auditory pathways (see Figure 2). The hair cells on the basilar membrane project to the spiral ganglion cells. Neural signals travel from the ganglion through the auditory nerve to the cochlear nucleus, to the superior olivary complex and through the lateral lemniscus to the inferior colliculus, then continuing through the medial geniculate nucleus of the thalamus until reaching the primary acoustic cortex.

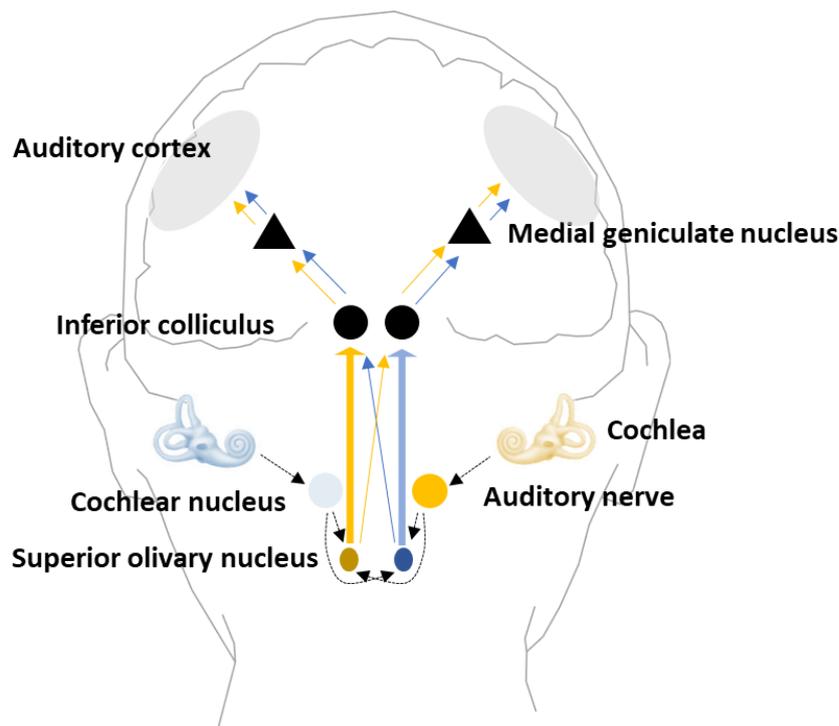


Figure 2 SCHEMATIC OF HUMAN AUDITORY SYSTEM

Sound source travels through the ear canal and turning into neural signals that reach the auditory cortex. Picture inspired and adapted from Henning, Sabic, & Hout, 2018; Shepherd, Seligman, & Fallon, 2015

A first encoding of the binaural auditory cues occurs at the level of the neurons of the superior olivary nuclei. ILDs are encoded in the lateral superior olivary nuclei, which are inhibited by sounds in the contralateral ear and excited by sounds in the ipsilateral ear. ITDs are encoded in the medial superior olive nuclei. The first encoding of the monaural spectral auditory cues occurs instead in the dorsal cochlear nuclei (Bear et al., 2020).

At all these stages, the tonotopic organization present in the cochlea is retained. To date, the only structure of the mammal brain in which neurons follow a topography that represents the location of sounds in space has been found outside the primary afferent auditory pathway, in the superior colliculus. In this subcortical structure, visual and auditory space are aligned and mutually adjust, as reported by King and colleagues in a neurophysiological study on adult ferrets (King, Hutchings, Moore, & Blakemore, 1988). The superior colliculus is a laminated structure: the dorsal layers receive visual input directly from the

retina, neurons in the deep layers respond to auditory, tactile and visual stimuli. The rostral part responds selectively to visual and auditory stimuli in the front, whereas the caudal part responds to the ones presented to the back of the body. This structure represents the first acoustic map of the space, but at the same time also it is dedicated to the integration of inputs coming from different sensory modalities because it contains spatial maps for each sense which are aligned and connected. King proposed two possible advantages of this laminated structure. First, it facilitates multisensory integration and consequently it favours events localization. Second, it allows to easily transform multisensory inputs into adaptive motor actions, such as the orienting of the eyes, head and body. Thus, the superior colliculus plays an active role in the transformation of the stimuli coming from the various senses into aligned maps that serve to control appropriate orienting responses to the external stimuli (King, 2004). The organization of this subcortical structure corroborates the idea of studying acoustic space as a multisensory and motor construct (2.1.2 and 2.1.3).

The combination and integration of information at the subcortical level is not sufficient to obtain a stable representation of the acoustic space. This representation is formed at the cortical level (van der Heijden, Rauschecker, de Gelder, & Formisano, 2019). The first region of the cerebral cortex implicated in acoustic perception is the primary acoustic cortex (PAC) in the temporal lobe, which receives afferents from the medial geniculate nucleus through acoustic radiation (van der Heijden, Rauschecker, Formisano, Valente, & de Gelder, 2018). This primary area is interconnected to the secondary auditory cortex and to multiple higher-order areas in the temporoparietal and inferior frontal cortices constituting a network (Tardif, Murray, Meylan, Spierer, & Clarke, 2006). The neurons of the primary auditory cortex respond as a function of frequencies (tonotopic organization), but some are also tuned to sound intensity. Furthermore, some neurons in the primary auditory cortex have been observed to respond selectively when a sound reaches both ears at the same time, showing sensitivity to binaural interactions (Bear et al., 2020).

In this wide-spread network of sound localization processing a crucial role is played by the posterior auditory cortex: the planum temporale (Deouell, Heller, Malach, D'Esposito, & Knight, 2007; Van der Zwaag, Gentile, Gruetter, Spierer, & Clarke, 2011; Zimmer & Macaluso, 2005). Brunetti and colleagues tested participants using a passive listening protocol (Brunetti et al., 2005). They used both fMRI and MEG to investigate the cortical areas activated during processing of sounds coming from various positions in space and the temporal dynamics of the sound localization process. These authors have shown the complex neural circuit implicated in this task, which comprised the primary auditory cortex in Heschl's gyrus, the superior temporal gyrus, the supramarginal gyrus in the inferior parietal lobe, and the inferior and middle frontal lobe. Furthermore, parietal and frontal areas such as the inferior parietal cortex (Krumbholz et al., 2005) and the supramarginal gyrus are involved in this network. The path that emerged from MEG data revealed the sequence of processing: Heschl's gyrus responded 139 milliseconds after

the auditory stimulus, the superior temporal gyrus at 156 milliseconds and the inferior parietal lobule and the supramarginal gyrus responded maximally at 162 milliseconds. These results emphasized the idea of a spatial and temporal progression involved in sound localization, which progressively recruits a widespread dorsal network involved in spatial processing and orienting of attention. Furthermore, they demonstrate the involvement of both hemispheres in processing auditory space, but with a stronger involvement of the right hemisphere (as previously suggested in other works: Lewald, Foltys, & Töpper, 2002; Spierer, Bellmann-Thiran, Maeder, Murray, & Clarke, 2009).

The role of the parietal lobe leads to consider the well-known two-streams hypothesis because of its involvement in the classical dorsal “where path” of the visual system (Goodale & Milner, 1992). The concept of dual processing streams has been adopted to study acoustic perception (Lomber & Malhotra, 2008; Rauschecker & Tian, 2000), yet it is still debated whether this is a too simplistic explanation. Some researchers proposed more complex models considering various sub-streams and multiple branches (Kravitz, Saleem, Baker, & Mishkin, 2011) and more complex involvement of the primary auditory cortex. The primary auditory cortex (PAC) is involved in spatial auditory processing during active, goal-oriented localization and its activation is modulated by the task (see van der Heijden et al., 2019 for a review). Nonetheless, the concept of dual processing streams is still actual, as suggested by Rauschecker in his review published in *Cortex* in 2018 (Rauschecker, 2018).

In the framework of the dual-stream model, sound processing passes through earlier stages of sensory processing in the primary acoustic cortex and goes up to higher-order level areas splitting in two different functionally specialized pathways. A ventral pathway processing sound object identity (the what) and a dorsal pathway dedicated to spatial processing (the where). The dorsal spatial pathway projects to the inferior parietal lobule (IPL) and to the superior frontal sulcus (see the meta-analysis by Arnott, Binns, Grady, & Alain, 2004) and it comprises a network involving the premotor cortex and the dorso-lateral pre-frontal cortex. These areas are involved in auditory motion processing (Poirier et al., 2017), in temporal processing (Schubotz, Von Cramon, & Lohmann, 2003) and in sensorimotor function (Rauschecker, 2018).

Rauschecker (2018) describes the dorsal path as strongly related to the sensorimotor integration and feedback control systems. In this renovated dual-stream model of auditory processing, the dorsal pathway is implied in the creation of an internal model of the outside world comprising the body. This review pointed out the importance of considering the dorsal pathway beyond the classic “where” system. On the contrary, this pathway combines sensory-motor elements to organize and produce behaviour (Rauschecker, 2018). In this new proposal, the authors have integrated various perspectives to redefine the dorsal path. As discussed by Miller and Goodale (2008) for vision, the dorsal stream processes the moment-to-movement kinematics of the action, rather than more static visual representations of object. Similarly, Rauschecker

suggested that the auditory dorsal pathway should be conceived as more related to “how” we interact with objects and thus the actions we direct towards them. This theoretical model is particularly interesting in relation to the reaching-to-sound approach that I have employed throughout this thesis (see Study 3, 4 and 5).

In line with the idea of a more active and motor role of the sound localization process, a recent review on the topic (van der Heijden et al., 2019) has considered the spatial hearing network during more active sound localization tasks. In addition to this marked motor component, active sound localization is also linked to the orientation of attention. Allocating attention to specific locations speeds up processing of sound localization (Johnen, Wagner, & Gaese, 2001) and facilitates speech understanding in noise (Oberfeld & Klöckner-Nowotny, 2016). This effect is the result of top-down mechanisms which can influence the activity of the primary auditory cortex. There are connections from higher-order areas in the brain as the dPFC to PAC (named direct feedback route) or the information can be mediated by the planum temporale (indirect feedback route) (Figure 3).

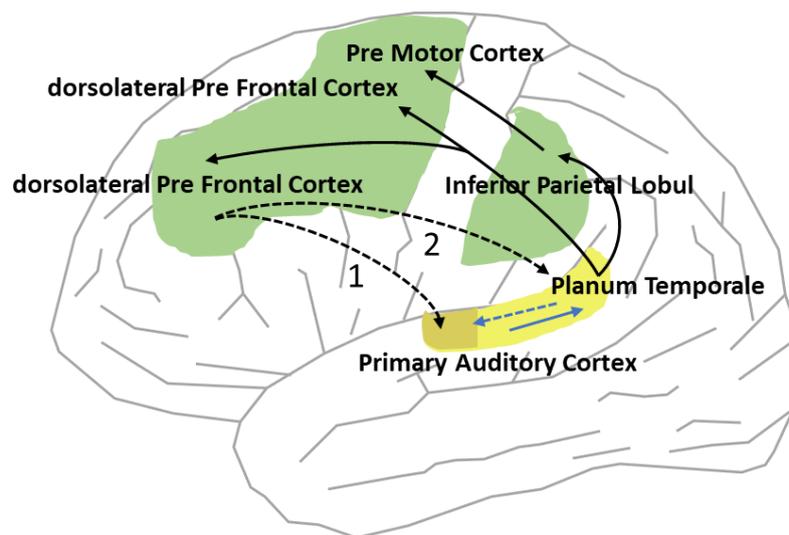


Figure 3 CORTICAL NETWORK OF ACTIVE, GOAL-ORIENTED SOUND LOCALIZATION PROCESSING

Auditory cortical regions were drawn in yellow, with lighter yellow representing increased spatial sensitivity. The green areas are sensorimotor regions of the auditory dorsal stream. Functional cortical connections are represented by the solid arrows, while the dashed arrows indicate feedback connections. As suggested by van der Heijden et al., (2019), the arrow 1 indicates a direct feedback path for top-down modulation of the PAC activity by the dorsolateral prefrontal cortex. The arrow 2 indicates the feedback route mediated by the planum temporale. Picture and description adapted from van der Heijden, Rauschecker, de Gelder, & Formisano, 2019.

The here presented evidence emphasizes the complexity of the neurocognitive process subtending sound localization. All agree in considering sound localization an active process of our brain resulting from a combination of input signals (such as auditory cues, but also coming from other modalities). The connections of the structures underlying the localization of sounds with the information of other sensory modalities are proof of the central role of multisensoriality when localizing sounds. The incoming input signals are integrated with more top-down mechanisms, in which action planning and task relevance play

important roles. These top-down modulation paths emphasize the importance of considering sound localization as a proper goal-oriented action.

1.1.3. Sound localization in a visual world

While spatial hearing relies on the functionality of the auditory system, it co-evolved with the other sensory systems and particularly with vision. Heffner (1997) has argued that “a primary function of audition is to direct visual attention for scrutiny of sound sources. This function seems to have been a major source of selective pressure affecting the evolution of sound localization among mammals” (p 51, Heffner, 1997). Evidence in this direction is the strong correlation between sound localization abilities and width of field of best vision (Heffner, 1997). Species with narrow fields of best vision are very accurate in sound localization, whereas species with larger best field of view appear to rely less on accurate spatial hearing as they have lower sound localization thresholds (i.e., good sound localization abilities compared to other mammals). By contrast, cattles possess a field of best view of great width paired with high sound localization thresholds (i.e., poor sound localization). This correlation suggests that visual and auditory system may have co-evolved in supporting the ability to detect and orient to distal stimuli in the environment. By allowing sampling of the environment all around the animal, spatial hearing may have co-evolved as a mechanism for bringing novel and relevant events into the field of view.

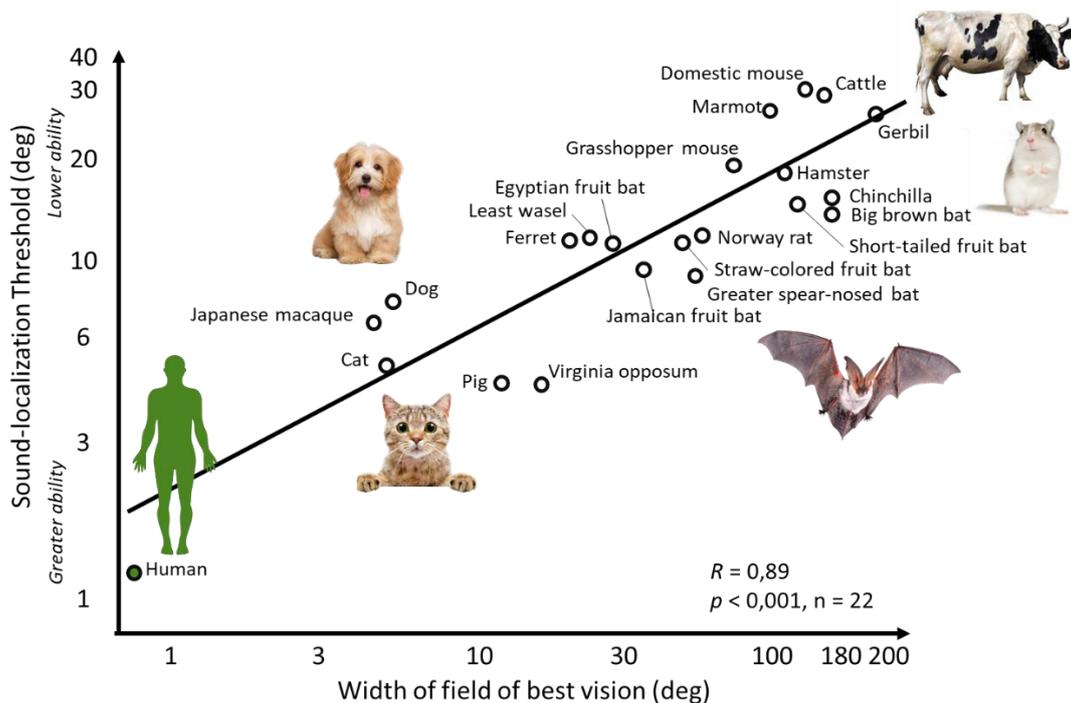


Figure 4 CORRELATION BETWEEN SOUND LOCALIZATION THRESHOLD AND WIDTH OF FIELD OF BEST VISION
Species with narrow fields of best vision are very accurate in sound localization (lower values on y axis), whereas species with larger best field of view rely less on accurate spatial hearing. Picture inspired and adapted from Heffner, 2018 and Heffener, 1997.

Examples of the close link between spatial hearing and spatial vision can be found in the literature on animal models. In a pioneering work in 1985, Knudsen and Knudsen took advantage of the use of prisms in the context of auditory spatial mapping recalibration. They demonstrated that barn owls raised with one ear plugged for 41-96 days made systematic errors in auditory localization, even after earplug removal (Knudsen & Knudsen, 1985). When monaurally occluded, barn owls were trained to orient their heads toward sounds and visual stimuli presented. The last week of monaural stimulation authors mounted in front of a group of animals' eyes a prism, while two animals were left with normal vision and two others were blindfolded using eye occluders. The two animals in vision unaltered condition were able to correct sound localization bias after few weeks. Conversely, the two deprived of vision did not correct it. The animals which wore prism adjusted auditory space perception to match the visual stimulation experienced through the prism. The results demonstrated the strong link between vision and sound localization and the importance of visual information to provide spatial reference on which to build the spatialization of sound. The tuning between visual and acoustic information assumes an important role during the processes of recalibration and thus it must be considered particularly in the altered listening conditions (see 2.2.). On the same line of investigation, another pioneering study on ferrets (King et al., 1988) studied the effect of deviating one eye laterally determined by section of the medial rectus muscle on the auditory map on the superior colliculus. They found that a small displacement of visual topography determined a compensatory shift in the auditory map, even if not exceeding a certain degree of alteration, which instead resulted in a disorganization of the auditory map. This strong link between vision and sound localization is also supported by the fact that people localize better the sound in the frontal space. As pointed out by Akeroyd and other authors, the frontal space is the visible space and thus, sound localization may favor visual attention (Heffner & Heffner 1992 and Hafter & de Maio, 1975; Pumphrey, 1950 in Akeroyd, 2014).

Other examples that support this relationship come from studies on audio-visual illusions. In the classical ventriloquist paradigm, the illusory effect is based on the idea that people perceive the sound as if it originates from the mouth of the puppet (visual source) instead of the mouth of the ventriloquist (actual auditory source; Alais & Burr, 2004; Bertelson & Aschersleben, 1998; Driver & Spence, 2000). This happens because the brain collapses together the information from both auditory and visual modalities and creates a unique percept based on spatial proximity and temporal congruency. Putting together these cues, the most natural position to locate the source is the mouth which is moving: the puppet's one. Interestingly, vision captures sound localization especially when sounds are more difficult to localize, but sound capture vision when this information is blurred (i.e., it is made less reliable in space; Alais & Burr, 2006). Thus, the brain can calibrate the association between auditory cues and spatial location in external space, exploiting the more spatially reliable information available. Researches in humans have exploited this illusory principle. Some authors exposed participants to synchronous but spatially discordant auditory

and visual stimuli (as in the ventriloquist phenomenon). The exposure determined a recalibration in sound localization resulting in an after-effect: when they were asked to localize sounds, participants judged their spatial position as shifted towards space where associated visual stimuli had been previously presented. Frissen and colleagues (2011) demonstrated that the recalibration is extremely fast and strong (Frissen, Vroomen, & de Gelder, 2012). The results strongly supported the role of visual inputs as essential when studying sound localization.

Despite the evidence supporting this link and the idea that visual cues from the environment can be exploited to locate the auditory stimulus using a visual spatial frame of reference (Warren, 1970), studies that have measured people's abilities to locate sounds have often underestimated the presence of visual cues in their setting. Some studies allowed full vision of the loudspeakers (e.g., van Hoesel et Tyler 2003), whereas in others, although they hid loudspeakers behind visible barriers, it was possible to see the whole room (e.g. Nava, Bottari, Bonfioli, Beltrame, & Pavani, 2009; Pavani, Farnè, & Làdavvas, 2003; Pavani, Meneghello, & Làdavvas, 2001). A notable exception is a study from the 1980s by Shelton and Searle (1980). In this study, they manipulated the possibility to have visual information about sound sources and showed that when a speaker array was placed directly in front of the subjects, vision of the sources facilitated accuracy compared to a blindfolded condition – even when the exact position of the target sound remained unknown (Shelton & Searle, 1980). This study represents the first attempt to manipulate directly the effect of visual scenarios during acoustic space perception. Despite it is intuitive to think that having a certain degree of visual information on the sources facilitates localization (e.g. seeing the speaker array), the role of the visual scene - beyond the possible sound sources- and its complexity in the acoustic space analysis, are still unclear. This experimental question that emerged in the 1980s can be addressed today through new methodologies. More recent works (more detailed in the introduction to Chapter 3, Ahrens et a., 2019; Majdak, Goupell, & Laback, 2010) took advantage from virtual reality to control visual stimulation independently from the acoustic one. Also, in the present thesis, I proposed a study exploiting this cutting-edge VR technology to obtain new and more precise answers regarding the role of the visual scene when localizing sounds.

Considering Warren's idea and the evidence described in the previous paragraphs, it may be argued that visual information alters and shapes acoustic space perception. To what extent this visual information is useful to us when localizing sounds? Do visual cues all have the same relevance for sound localization? Is only the position of the source involved or also the geometry of the entire space? Does seeing an empty visual environment impact on spatial hearing?

I will address this question in the first experimental contribution: Study 1 “The impact of a visual spatial frame on real sound-source localization in virtual reality”.

1.1.4. Head movements and spatial hearing

Whenever relevant sounds are heard in the environment, humans as well as other hearing animals react with a typical behavior: they turn their head toward the sound source. Head orienting to sounds is a behavior detected since birth. Infants in the first 25 hours from birth already exhibit head orienting toward sounds (Clifton, Morrongiello, Kulig, & Dowd, 1981). This behavior has been documented also in infants a few days old (Muir, Humphrey, and Humphrey, 1994) and in infants a few months old (Field, Muir, Pilon, Sinclair, & Dodwell, 1980; van der Meer, Ramstad, & van der Weel, 2008). Head orienting is a spontaneous movement that people perform to face the speaker also in adulthood. Under normal hearing conditions, people try to stand in front of the speaker, both when they localize sounds or when they were listening in noise (Iwaya, Suzuki, & Kimura, 2003; Thurlow & Runge, 1967; Valzolgher, Alzhaler, et al., 2020). Further evidence supporting the idea that head movements are spontaneous reactions to sounds is that the cognitive system weights auditory cues considering also the body position and head rotation (Blauert, 1997; Middlebrooks & Green, 1991).

Head movements are not only spontaneous behaviors, they can be implemented intentionally. During a sound localization task, participants can move the head to change auditory cues and obtain a better listening condition. Rotating the head around the vertical axis impacts on ITD and ILD, while tilting the head primarily modifies monaural cues (Kato, Uematsu, Kashino, & Hirahara, 2003; Perrett & Noble, 1997b). Thus, the dynamic content of binaural and monaural cues resulting from the displacement of the head can lead to performance benefits. The dynamism of the auditory cues helps sound localization because it enriches the acoustic experience. Moving the head means changing the auditory cues corresponding to a given point of the space occupied by the sound source. Increasing these cues leads to a greater chance of extracting the correct spatial position. Several studies have suggested that head movements promote sound localization in horizontal and vertical dimensions (Morikawa & Hirahara, 2013; Perrett & Noble, 1997a, 1997b; Thurlow & Runge, 1967) and reduces front/back confusion errors (Iwaya et al., 2003; Kim, Barnett-Cowan, & Macpherson, 2013; Wallach, 1940; Wightman & Kistler, 1999).

Although the importance of head-movements when sound localizing has been emphasized since 1940 (Wallach, 1940), systematic investigations on their role in spatial hearing started only in the last decades. This may reflect the lack of technologies and advanced methodological strategies which may permit to monitor the position of the head in real-time. The first researchers who tried to analyze the role of active movements during listening were Thurlow and Runge in 1967. They have investigated head movements by moving the participants' head during sound presentation using a custom-made head-moving equipment (Thurlow & Runge, 1967). The authors documented the effectiveness of induced head rotation in reduction horizontal and in vertical localization error. Afterwards, Perrett and Noble (1997)

demonstrated that letting normal hearing participants free to perform head-movements while hearing can improve their ability to localize sounds in the horizontal and vertical dimensions as compared to motionless condition (Perrett & Noble, 1997b, 1997a). During the head movements condition, participants were instructed to rotate the head rightward and leftward. Thus, their head-movements were guided and controlled using visual cues and there were not completely spontaneous. Yet, they did not measure precisely the amplitude of each movement performed. Much evidence also supports the critical role of head movements in resolving the front/back confusion. Iwaya and colleagues (2003) tested front-back discrimination comparing two different listening conditions. During the “restricted condition”, participants were instructed to keep their head immobile, whereas during the “free condition” they were free to move their head naturally (Iwaya et al., 2003). The results of this work confirmed the benefit of head movements and found that the front-back error rate was reduced by head movements, especially when the sound duration is long (2 seconds instead of 0.5 s). Methodologically, in this study they measured head movements using a magnetic sensor placed at top of the head. The authors found that during the free listening condition, head movements were aimed at positioning the sound frontally.

While these findings agree in attributing a prominent role to head movements during acoustic space perception, they raise an interesting question regarding the differences in degree of the intentionality of the movements implemented in the active/free conditions. In this sense, movements cannot be considered all the same. In the first example the movements were passive (Thurlow & Runge, 1967), in the second examples more active (Perret and Noble 1997; Iwaya et al., 2013). Spontaneous and intentional movements can play different roles and reflect different mechanisms. A first step to investigate this aspect was taken by a study by Brimijoin and Akeroyd (2014), who tested whether self-motion is processed differently from source motion (Brimijoin & Akeroyd, 2014). Particularly, they asked participants to judge whether a female voice speech was to the left or the right of a male voice speech, as in the classical paradigm to measure minimum audible angle (MAA). The task was performed in a “self-motion” condition in which participants turned their heads around the vertical axis and in a “source-motion” condition in which listeners remained still, and the sound locations were moved virtually recreating the trajectory of the “self-motion” condition. With this methodological solution, the authors were able to disentangle the contribution of self-motion without impacting the motion with respect to the head and found that minimum audible angle was smaller during the self-motion compared to the source-motion condition. This result contributed to deepen the role of head movement in acoustic space perception and it highlighted the link between the spatial auditory cues and self-motion cues.

Evidence in the literature, therefore, agrees that active self-promoted head movements are considered crucial when locating sound sources. However, the implications of this active behavior are still little investigated, in normal hearing and – especially – in individuals with hearing difficulties. What are the

elements that favor head movements when listening with the aim to localize sounds? When are they elicited or, viceversa, when they are no longer implemented? In which context are the movements implemented or limited?

The theme of head movements is recurrent in the empirical studies of this doctoral dissertation. I have focused on investigating to what extent moving the head can help sound localization when listening is impaired and what role it may play in relearning spatial hearing abilities. Furthermore, as suggested by Brimijoin, McShefferty and Akeroyd in 2012, head movements can also be implemented to overcome a difficult situation. Testing listeners with asymmetry in the hearing threshold between their left and right ears (>16 dB) in a hearing in noise task, they observed that head movements led participants to reach the most favorable listening posture (Brimijoin, McShefferty, & Akeroyd, 2012). By turning their heads, participants increased the level of the signal in their better ear (Brimijoin & Akeroyd, 2016). The authors judged this behavior as a simple, easily implemented strategy. This example, although concerning a task not directly linked to spatial listening, indicates that head movements can also be used actively as a self-regulation strategy to improve hearing abilities. In the experimental contributions proposed here, I have examined if similar adaptation strategies can also be implemented to solve auditory spatial tasks (Brimijoin & Akeroyd, 2016).

I have addressed this issue in experimental contribution: Study 2 “Spontaneous head-movements improve sound localization in ageing adults with presbycusis” and in the studies Study 3 “Adapting to altered auditory cues: generalization from manual reaching to head pointing” and 4 “Reaching to sounds improves spatial hearing in bilateral cochlear implant users - A Randomized Clinical Trial”.

1.2. Re-learning sound localization

Spatial hearing is plastic and can adapt to different contexts. We can re-learn to localize sounds through the course of our life. Our cognitive system can adapt to changes in our body, such as, the growth of the head which affects the correspondences between auditory cues and spatial directions (Clifton et al., 1988). Compensatory adjustments that occur in the auditory space map in the superior colliculus have been documented during the early stage of life. As discussed by King and colleagues (King, Parsons, & Moore, 2000) this may suggest that plasticity may be stronger during infancy to respond effectively to these changes. However, evidence in the literature has demonstrated that our brain can adapt to altered auditory cues even later, during adulthood. This plasticity could result from changes occurring in subcortical structures, such as the superior colliculus, as well as cortical structures, such as the parietal cortex (King, Parsons & Moore, 2000).

The study that first investigated if humans were capable of relearning the ability to localize sounds was conducted by Hofman and colleagues (1998). These authors demonstrated that adults can accommodate

to modified spectral cues. They tested sound localization abilities of four normal hearing participants listening with bilateral ear molds. Wearing ear molds altered the possibility of using monaural cues and, as a consequence, the listeners' performance in the vertical plane decreased dramatically. Critically, after a variable period of altered listening experience, all participants improved spontaneously their sound localization skills in elevation. This finding constitutes the first evidence that adults can relearn to localize sounds in altered listening conditions.

How can the auditory system adapt to altered acoustic information and what are the mechanisms involved? In altered listening situations, like the one proposed by Hofman, the cognitive system must relearn to localize sounds with the newly available auditory cues. In the literature, two main mechanisms have been identified to support this re-learning phenomenon: auditory cues reweighting and auditory cues remapping. Below we illustrate each mechanism in turn, but it is important to note that they are not mutually exclusive (Hofman, Van Riswick, & Van Opstal, 1998).

The term auditory cues reweighting has been introduced by Van Wanrooij and Van Opstal in 2007 to describe the notion that the auditory system can learn to cope with altered auditory cues by changing how each cue (monaural or binaural) contributes to the resulting percept of sound location (Van Wanrooij & Van Opstal, 2007). The underlying assumption is that, under normal listening conditions, a variety of auditory cues contribute to sound localization. In the mechanisms that creates a correspondence between the auditory cues and the locations in external space the brain integrates the various auditory cues, by weighting each as a function of its reliability (Keating & King, 2015).

Van Wanrooij and Van Opstal tested participants in a monaural listening situation and found a mis-localization bias toward the unplugged side. In addition to plugging one ear, they minimized the possibility of using binaural cues by presenting low intensity sounds. Indeed, at low intensity the effect of the ear plug is maximum and thus participants experienced a real monaural listening. Conversely, at higher intensities, they still received partial stimulation through the ear plug and thus it was not a real monaural listening. Participants performance at their unplugged side was better for low-intensity stimuli. Thus, these findings revealed that spectral cues of the unplugged ear were involved in sound localization in the horizontal plane and demonstrated that when binaural cues are minimal and ambiguous as for low-intensity stimuli, the spectral shape cues of the unplugged ear become more relevant.

They also tested the ability of participants to judge sounds' positions in elevation in the same context. They reported an increase of errors, even if smaller than the effect on azimuth. Starting from these results, the authors hypothesized a model of sound localization in altered hearing situation. The greater the reliability of the cues, the more relevant they are in the weighting process. When judging the horizontal position of a certain source in monaural hearing condition, both the binaural cues and the spectral cues

of the plugged ear become uncertain and poor. Thus, monaural cues of the unplugged ear gain importance, but they are not able alone to extract horizontal spatial information resulting in a mis-localization of sounds toward the unplugged ear. Similarly, the results on the vertical dimension suggest that binaural cues and monaural cues of both ears are weighted together with the monaural cues when judging this dimension.

The process of cues reweighting has been observed in ferrets by Keating and colleagues in 2013. When plugged monaurally, ferrets reweighted spatial cues by relying on the unaffected spectral cues. The reweighting process was also supported by changes in neural responses activity of the primary auditory cortex. The authors measured the spike rates of high-frequency neurons in the primary auditory cortex and found that their sensitiveness was modulated by hearing experience. When they were plugged monaurally, the neuronal responses became more sensitive to the monaural spatial cues provided to the unplugged ear and less to the plugged one. If plugging intermittently (i.e. interspersed by periods of normal hearing), ferrets were able to reweight the spatial cues selectively during monaural listening, while reweighting process disappeared during normal hearing. In this study, the authors confirmed, on one hand, the possibility of the auditory system to reweight the cues as a function of the most reliable cues available. On the other hand, they demonstrated the auditory system's ability to use reweighting strategy as a function of the specific acoustic condition. This observation highlights that auditory system can maintain two distinct models/maps of the auditory space which can be used as a function of the auditory context (Keating et al., 2013).

Another process involved when re-learning spatial hearing is auditory cue remapping. The term cues remapping is linked to the longer process of relearning and corroborates the hypothesis that human auditory system can create multiple correspondences between specific cues and coordinates in the external space. Remapping means being able to create through experience new correspondences between cues and spatial positions, which remain available during life.

An interesting study revealing this process is the work of Trapeau and Schönwiesner (Trapeau & Schönwiesner, 2015). These authors tested participants using digital bilateral ear-plugging, which artificially shifted the natural ITDs of participants, for seven consecutive days. After 4 days, participants tested in a sound localization task showed improvement, which was supported by functional MRI data. At the cortical level, they observed a different neural tuning of the new ITDs. Specifically, they measured the changes in the hemispherical lateralization of the cortical activity as index of horizontal localization coding. The behavioral adaption to ITDs changes was accompanied by modification of cortical activation. The authors also documented that the new sound-space correspondence did not interfere with the old one. Participants tested after removing the plug, did not show any after-effect, revealing that a new independent acoustic map has been created. Disrupting monaural cues instead of binaural ones, in the

presented above work of Hofman and colleagues, authors showed that participants, after removal of the molds, performed comparably as before molding. Notably, participants were able to localize in both listening conditions. This finding demonstrates that the new acoustic map created was added and not overwritten to the pre-existing one (Hofman et al., 1998).

A further study on the topic comes from animal literature, Keating and colleagues in 2015 tested ferrets in a horizontal sound localization task. Specifically, they observed that normal hearing ferrets performed accurately the task, while their error increased when their left ear was plugged. Differently, smaller errors were found when juvenile-plugged ferrets were tested (which have been monaurally plugged from hearing onset). The authors recorded also neuron activity in the primary auditory cortex to measure interaural level difference sensitivity in each ferret group. In virtual one-ear plugged condition, while the normal hearing ferrets' neurons sensitivity declined significantly, juvenile plugged ferrets retained sensitivity to ILD which appeared to be shifted toward the unplugged ear. Under normal hearing, neurons in both hemispheres preferred stimuli whose levels were higher in the opposite ear. This response pattern was affected when ear plug was used in the control group. Differently, for juvenile plugged ferret neurons in both hemispheres preferred stimuli from the unplugged ear. This study has created a hearing condition in which spectral cues are compromised and the abnormal binaural cues are the only available and reliable to be used. In this case, it was impossible for the acoustic system to rely on residual cues and to ignore or partially consider the altered ones in a reweighting process. Conversely, the strategy adopted is much more related to a remapping process, which, precisely requires a longer period to be implemented as in the juvenile-plugged ferret.

As suggested in this work of Keating and colleagues in 2015, cue remapping and cue reweighting highlight the flexibility of acoustic space processing and may occur at different stages of auditory space relearning processes (Keating, Dahmen, & King, 2015). As discussed by Keating and colleagues in 2016, these processes are not mutually exclusive and can occur concomitantly. However, it is still unclear how these two processes are integrated and combined. A hypothesis is that while the process of cues reweighting can be considered as an immediate process happening suddenly, and that does not determine by itself the creation of a new acoustic map, cue remapping subtends a longer process of creation and storage of new auditory maps. Thus, cues reweighting could be considered the first step toward the recalibration of the acoustic space, while cues remapping is responsible for the consolidation of multiple auditory maps (Keating, Rosenior-Patten, Dahmen, Bell, & King, 2016).

Having documented that it is possible to weigh the cues according to different auditory contexts and to create new acoustic maps, it may be asked whether these processes can be fostered or accelerated. In the last decades, researchers have demonstrated the possibility to update auditory cues during the lifetime both testing animals or artificially modified human listening (i.e. altered listening models). To alter

binaural cues, researchers have often plugged one ear using earplug and muff and caused a change in the sound level reaching the two ears, affecting particularly ILDs, but also ITDs (Keating & King, 2013; Strelnikov, 2011). To study the perception of acoustic space when monaural cues are altered, two main methods have been adopted. The first method consists of inserting in the concha cavity of the outer ear an ear-mold. The aim is to change the anatomical conformation of the pinnae and thus modifying spectral cues (Hofman et al., 1998). The second approach directly manipulates the sound waves reaching the ear canal. It is possible to register sound coming from various spatial positions by inserting a microphone in the ear canal. This technique gives the possibility to capture the effect of ear shape on sound structure creating a virtual spatialization of sound. These stimuli could be delivered through headphones and are named head related transfer functions (HRTFs). When HRTFs are not individualized (they are not created using the head of the listener) the experience of listening is distorted. People reported listening as with “the ears of a stranger” and sound localization is distorted, especially in elevation, the dimension that most relies on monaural cues.

As already mentioned at the beginning of the chapter, multisensory stimulation can influence sound localization. Thus, it can be used also to train the acoustic space perception. For instance, Strelnikov and colleagues in 2011 tested the effect of an audio-visual training and compared it to an auditory-only training (Strelnikov, Rosito, & Barone, 2011). The performance improvements in the sound localization task were larger after the multisensory stimulation-based training compared to the other. This finding was supported by various evidence proposing also training based on multisensory feedback on humans (Shinn-Cunningham, Durlach, & Held, 1998; Zahorik, Bangayan, Sundareswaran, Wang, & Tam, 2006) and ferrets (Isaiah, Vongpaisal, King, & Hartley, 2014). These studies demonstrated the crucial role of multisensory stimulation and visual feedback to favor relearning.

Head movements are fundamental when listening, but they were often limited by some practical issues related to the experimental setting (e.g. chinrest). Carlile and colleagues (2014) overcame this limit and tested the contribute of active listening as compared to more static situations. More precisely, they tested the effect of four different training paradigms in accelerating the processes of adaptation to altered spectral cues. Participants were invited to wear small ear molds and they were tested in a nose pointing to sound task before and after ten days of accommodation. During the audio-visual sensory motor paradigm, participants were encouraged to explore the space around the target by moving their heads. Notably, the greater extent of accommodation was observed in this audio-visual sensory motor paradigm as compared to the ones proposing only a visual feedback or not presenting any feedback (Carlile, 2014).

A recent element that authors have started to consider in the context of acoustic space training is the hand motor action towards sounds. Some studies provided initial evidence in support of a key role of dynamic and sensorimotor interactions with sounds in fostering adaptation to altered auditory cues

(Mendonça, Campos, Dias, & Santos, 2013; Parseihian & Katz, 2012; Valzolgher, Campus, Rabini, Gori, & Pavani, 2020). The effective paradigms proposed in these studies comprised motor interaction by involving participants in moving their hand or by implementing reaching to sound sources action in gamified and active context (Honda et al., 2007; Ohuchi, Iwaya, Suzuki, & Munekata, 2005; Parseihian & Katz, 2012). Despite the interesting and attractively applied scenarios created, none of these studies tested whether reaching movements are distinctive in promoting the useful strategies for adapting to altered spatial hearing conditions.

Most of these studies still have methodological limitations. First, as suggested by Mendonca in her review (Mendonça, 2014), to test the effect of a training paradigm, it is fundamental to measure its generalization beyond the training task. It means that participants' abilities need to be assessed before and after participating in the target training. Importantly, the testing phase must differ from the training one to exclude any effects of habituation. Second, to effectively measure the effect of a certain training and to exclude habituation effects, the experimental design is fundamental. When proposing a between participants experimental design, a control group is needed. The control group performs both the testing phases, but without being trained in between. This solution gives the opportunity to monitor the improvement induced by the habituation to the testing situation and to disambiguate the spontaneous adjustment from the one caused by the training.

In the experimental contributions of this thesis, I placed emphasis on the aspects of active and motor interaction as a means of training studying more systematically its role in the relearning process. Using reaching action to explore the environment (even the acoustic space) emphasizes the idea that perception is itself an active process of exploration. This idea is very well conveyed by Gibson in the 60s, in one of his experiments on tactile perception demonstrating the perceptual advantage of a more active exploration (the "cookie cutter experiment") (Gibson, 1962). Consistent with Gibson's observations, Dehaene (2018), in his book "Apprendre!: les talents du cerveau, le défi des machines", highlighted the importance of active engaging in a task, as a crucial factor to favor relearning.

1.3. Toward a clinical model

The study of the mechanisms of spatial hearing does not respond only to theoretical questions, it is also relevant for accurate measuring of spatial hearing skills in people with impaired hearing and for studying how these abilities can be rehabilitated.

Hearing deficits are a major social, health, and economical challenge. Worldwide, there are about 466 million people with hearing problems. Besides children born with congenital hearing problems, approximately one-third of individuals over 65 years are affected by disabling hearing loss (WHO, 2020). In addition to affecting the possibility of listening to speech, hearing loss affects spatial hearing (Abel &

Hay, 1996; Dubno, Dirks, & Morgan, 1984; Eisenberg, 2007; Kumpik & King, 2019; Tasell, 1993). Hearing loss and deafness could be congenital or acquired during life. It may result from complications at birth, infectious diseases, chronic ear infections, use of oto-toxic drugs, exposure to excessive noise, and ageing (WHO, 2020). Based on the extent of hearing loss, they may be mild (20-40 dB), moderate (40-70 dB), severe (70-90 dB), or profound (>90 dB)¹ and it can affect one or both ears.

Hearing losses can also be classified as a function of the damaged region. Conductive hearing losses are caused by problems in the external or middle ear that prevents correct sound conduction. For instance, when the ossicles in the ear and eardrums do not work correctly because of infections. Conductive losses are mild or moderate and they are often temporary, but sometimes they are also permanent. In these cases, implants for middle ear or hearing aids can be useful. Sensorineural loss is usually permanent, and it is caused by the damage of the hair cells or the lack of them. They may be mild, moderate, severe or profound. Mild or severe sensorineural hearing loss can often be treated using hearing aids or middle ear implants, while cochlear implants are often a solution for severe or profound hearing loss. Furthermore, there are also mixed hearing losses, which are a combination of these two types of hearing loss and neural hearing losses. They are caused by damage or absence of the acoustic nerve and are generally deep and permanent. Neither hearing aids nor cochlear implants are helpful in this case since the acoustic nerve is unable to transmit enough auditory information to the brain. In some cases, a possible treatment is represented by brain stem implantation (ABI).

Given the variability of hearing losses, the way in which the ability to locate sound is altered also varies. This change depends on the degree of hearing loss, the frequency range it affects and the type of deficit. Performances change as a function of the auditory cues available. For instance, Noble and colleagues (Noble, Byrne, & Lepage, 1994) tested the localization ability of 87 bilaterally hearing-impaired listeners both in the horizontal and vertical planes, frontally and laterally. They reported that sound localization in the horizontal plane is little altered in neurosensory hearing loss, while it is compromised in people with conductive hearing loss. Concerning the localization on the vertical plan, it is affected in all types of hearing loss. Spatial hearing is altered also in people with unilateral hearing loss. Asymmetry of binaural cues causes difficulty in localization in the horizontal plane (Kumpik & King, 2019). It is important to notice that the possibility of exploiting the auditory cues depends on the sound's frequency. Thus, hearing loss could affect differently the threshold for each frequency causing differences in the way in which people can extract auditory cues and thus affecting sound localization differently.

¹ Note that classification criteria may vary slightly in different states. These data refer to the Italian system and are taken from: <https://www.burlo.trieste.it/audiologia-ori>.

Although this difficulty in perceiving the acoustic space is well-documented in the literature, sound localization is not considered in the clinical practice. Neither tests nor rehabilitation programs aimed at measuring and enhancing spatial acoustic abilities are proposed. The main intervention at present consists in using technological devices to restore the acoustic experience. Over the years, hearing technology increased its efficacy and provided benefit to hearing impaired people. The most popular devices, the hearing aids are worn by many people around the world and particularly the elderly population who suffer from presbycusis (Lunner, Rudner, & Rönnerberg, 2009; Tasell, 1993). However, hearing aids are not enough when hearing loss is too severe. In the last decades, it has become possible to bypass the damaged part of the hearing system by stimulating electrically directly the hearing nerve through a neuro-prosthesis, called cochlear implant. Mild or severe sensorineural hearing loss can often be treated using hearing aids or middle ear implants, while cochlear implants are often a solution for severe or profound hearing loss. As it will be described later, the use of these devices is not sufficient to effectively perceive auditory space and sound localization of people wearing these devices is still often compromised.

Given this context, theoretical contributions, accompanied by applied research on the effects of the acoustic space training paradigms tested directly on the clinical population, can contribute to give the appropriate emphasis to this spatial deficit, so far largely overlooked. In the experimental contributions of this thesis, I tested elderly patients with asymmetric presbycusis (Study 2) and patients with bilateral cochlear implantation (Study 5) in two distinct studies.

In the first case, I tested elderly with asymmetrical presbycusis, who represent an increasing share of people with hearing impairment. With increasing age, hearing capacity declines and, specifically for males, high frequencies are more difficult to identify (Gates, Schmid, Kujawa, Nam, & D'Agostino, 2000; Pearson et al., 1995). Identifying sound localization in elevation (Noble et al., 1994) and disambiguate front back confusion suffers more the effects of this acoustic impoverishment (Rakerd, Velde, & Hartmann, 1998). Furthermore, the presence of an asymmetry in the hearing threshold as the case I tested in Study 2, causes an impoverishment of auditory cues fundamental for localizing along the horizontal dimension. This population, therefore, represents a clear example in which the auditory cues to localize sounds are very impoverished, and yet they experienced typical development of spatial hearing skills during life (e.g., binaural integration, typical associations between auditory cues and space).

In the second case, people with severe to profound hearing loss can increasingly benefit of cochlear implants (CI), a neuroprosthesis that partially restore the hearing function (Moore & Shannon, 2009). Cochlear implants are systems composed by external and internal equipment. The visible external part comprises a microphone which picks up acoustic stimuli and transmits them to a sound processor placed usually behind the ear (Figure 5, step 1). This processor converts the acoustic input (vibration) into an

electric signal. Though a transmission coil the mechanical vibration is sent to an internal receiver. Both processor and receiver are kept aligned through a magnet (Figure 5, step 2 and 3). Then, the receiver transmits the electrical signal to the electrodes in the cochlea to stimulate directly the acoustic nerve (Krogmann & Khalili, 2020) (Figure 5, step 4). The interest in electrical stimulation of the acoustic nerve originates in the late 1950s, when Djourno and colleagues reported that using electric stimulation they succeeded in inducing hearing in two totally deaf patients. From that moment on, researches in Europe and in the United States led to an evolution of this technique to the point of making it a real medical device, with the first CI approved by the FDA for use in adults in 1996 and in children in 1997. To date, the number cochlear implanted users has increased exponentially. The technological progress and the widespread scholars' interest on this topic allowed people to benefit from highly efficient implants, especially in terms of speech performance (Zeng, 2004). Indeed, the cochlear implant is considered the gold standard in the treatment of profound and, in some cases, severe deafness in many European countries.

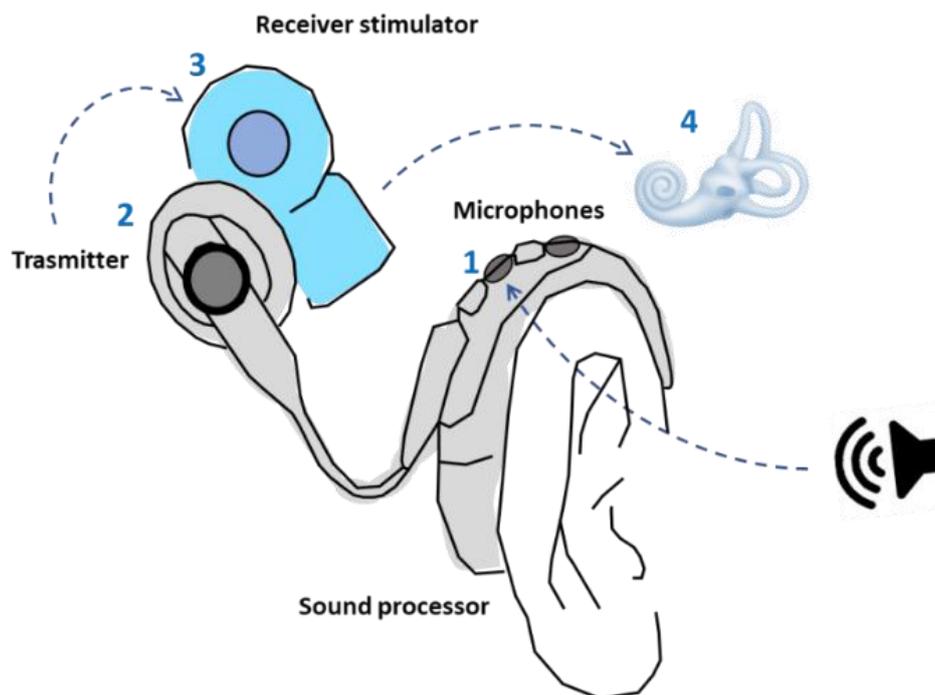


Figure 5 SCHEMATIC DESCRIPTION OF A COCHLEAR IMPLANT

(1) An external microphone picks up acoustic stimuli and transmits them to a sound processor placed usually behind the ear; (2) the sound processor converts the sound into an electric signal; (3) through a transmission coil the signal is sent to an internal receiver stimulator. (4) the receiver transmits the electrical signal to the electrodes in the cochlea

Although these devices restore the acoustic experience, they do not do so in full. To date, the acoustic space perception is still little considered. Researchers so far have mainly tried to make these devices more effective in capturing linguistic information. The information provided by the system (unilateral cochlear implant) or systems (bilateral cochlear implants) is not enough to effectively map the sound space as the normal hearing people (Grantham, Ashmead, Ricketts, Labadie, & Haynes, 2007; Jones, Kan, & Litovsky, 2014; Nava et al., 2009; van Hoesel & Tyler, 2003). Several factors impact on the

potential for rendering auditory space of this device. The most impacting factor concern the number of implants: having two implants compared to only one brings a great benefit because it restores a binaural experience of sound (Grantham et al., 2007; Grieco-Calub & Litovsky, 2010). In addition, some technical features of the implant such as the way in which the acoustic signal is encoded, the number of active electrodes, the type and number of microphones play a key role in speech recognition and may influence also spatial perception (Kokkinakis, Azimi, Hu, & Friedland, 2012; Wanna et al., 2014). Furthermore, the age of implantation, the initial degree of deafness, whether the person had experienced monaural or binaural acoustical information before, all influence the spatial hearing outcome (Kral, Dorman, & Wilson, 2019; Sharma, Dorman, & Spahr, 2002). Despite this great variability, the literature on the topic still demonstrates that bilateral cochlear implants users (BCI) do not perform as people with normal hearing, either in locating stationary sound sources (Jones et al., 2014; Kerber & Seeber, 2012), or when detecting the direction of moving sounds (Moua, Kan, Jones, Misurelli, & Litovsky, 2019). Thus, from a more applicative point of view, it is increasingly necessary to implement rehabilitation strategies that aim to strengthen the connection between acoustics cues and space in cochlear implant users. The objective of the latest experimental contribution (Study 5 – “Reaching to sounds improves spatial hearing in bilateral cochlear implant users - A Randomized Clinical Trial”) of the present thesis is answering to a more translational experimental research question: Can we train sound localization in bilateral cochlear implant users?

Chapter 2. **Study 1: The impact of a visual spatial frame on real sound-source localization in virtual reality**

This study has been published in *Current Research in Behavioral Sciences* in 2020. The following contents are reported from the same journal article.

Valzolgher, C., Alzhaler, M., Gessa, E., Todeschini, M., Nieto, P., Verdelet, G., Salemme, R., Gaveau, V., Marx, M., Truy., Barone, P., Farnè, A. & Pavani, F. (2020). The impact of a visual spatial frame on real sound-source localization in virtual reality. *Current Research in Behavioral Sciences*, 100003. <https://doi.org/10.1016/j.crbeha.2020.100003>

2.1. Abstract

Studies on audio-visual interactions in sound localization have primarily focused on the relations between the spatial position of sounds and their perceived visual source, as in the famous ventriloquist effect. Much less work has examined the effects on sound localization of seeing aspects of the visual environment. In this study, we took advantage of an innovative method for the study of spatial hearing – based on real sounds, virtual reality and real-time kinematic tracking – to examine the impact of a minimal visual spatial frame on sound localization. We tested sound localization in normal hearing participants (N=36) in two visual conditions: a uniform grey scene and a simple visual environment comprising only a grid. In both cases, no visual cues about the sound sources were provided. During and after sound emission, participants were free to move their head and eyes without restriction. We found that the presence of a visual spatial frame improved hand-pointing in elevation. In addition, it determined faster first-gaze movements to sounds. Our findings show that sound localization benefits from the presence of a minimal visual spatial frame and confirm the importance of combining kinematic tracking and virtual reality when aiming to reveal the multisensory and motor contributions to spatial-hearing abilities.

2.2. Introduction

In humans, as well as in other animals that can hear, the ability to localize sounds in space has evolved over the years within a multisensory environment. Under this ecological pressure, spatial hearing co-evolved with other sensory systems such as vision, which provides distal information about the environment (Heffner & Heffner, 2014; Heffner & Heffner, 1992). In addition, studies in animal and human models clearly showed that vision plays a critical role in the development of acoustic space perception (Hofman et al., 1998; Knudsen & Knudsen, 1985). When a listener is engaged in a sound localization task, there are at least two ways in which vision can contribute useful spatial information. First, vision can provide direct information about the auditory target, by revealing the position of the sound source in the environment (e.g., the listener hears and sees the bird tweeting on the tree). Second, vision can provide indirect information about the auditory targets, by revealing from which sector of space they may originate or by providing general information about the environmental spatial frame for encoding sound position (e.g., the listener cannot see the bird tweeting, but perceives the tree branches from which the stimulus originates).

The vast majority of studies that investigated audio-visual interactions in spatial hearing have been carried out in a context in which vision provides direct visual cues about sound position. In the typical experiment of this sort, the onset of the target sound is accompanied by a visual cue. When the visual information is veridical, listeners are more precise in sound localization compared to when no information is provided (Shelton & Searle, 1980; Tonelli, Brayda, & Gori, 2015). Instead, when the visual information is not

veridical, visual-capture of sound position typically emerges (the well-known 'ventriloquist effect') (Alais & Burr, 2004). For instance, Bolognini et al. (2007) demonstrated that veridical visual cues can enhance sound localization (Bolognini, Leor, Passamonti, Stein, & Làdavas, 2007). Participants sat in front of a plastic semicircular apparatus which comprised eight loudspeakers hidden behind a curtain. They were required to verbally judge sound location, reading aloud labels marking the position of each speaker. Crucially, the auditory stimulus was either presented alone or together with the visual stimulus. Results showed that spatially and temporally coincident visual stimuli improved sound localization accuracy. Consistent results have been observed also in infants (Morrongiello & Rocca, 1987). Moreover, these direct visual contributions to sound localization have been proved useful when training acoustic space perception (Rabini, Altobelli, & Pavani, 2019; Strelnikov et al., 2011; Valzolgher, Campus, et al., 2020).

Considerably less research has instead investigated indirect visual contributions to sound localization. Yet, the idea that vision of the surrounding environment can provide useful spatial information for spatial hearing dates back to the 1970s, when it was termed 'visual facilitation' (Warren, 1970). Warren was among the first to report that sound localization accuracy can improve when listeners localize unseen auditory targets with their eyes open than closed. In eye-open conditions, speakers were hidden from the subject using a fabric screen, but the overall environment was clearly visible. When interpreting the advantage observed in eye-open conditions, Warren proposed that participants use visual cues from the environment to place the auditory stimulus into a visual spatial representation, instead of using only an auditory frame of reference. In line with this early observation, a decade later Shelton and Searle (1980) showed that when a speaker array is placed directly in front of the subjects, vision of the sources can facilitate accuracy compared to a blindfolded condition – even when the exact position of the target sound remains unknown (Shelton & Searle, 1980). More recent works have replicated these pioneering studies, suggesting that seeing the environment can also enhance the precision of motor response and may thus affect participants' performance by facilitating their motor behavior. For instance, Redon and Hay (2005) showed that the presence of a visually structured background reduces pointing bias to visual targets (Redon & Hay, 2005). Interestingly, even the brief observation of the overall environment can improve spatial hearing (Tonelli et al., 2015).

Yet, in these classic studies as well as in more recent ones, it was difficult to disambiguate indirect contributions to sound localization that resulted from seeing the overall structure of the environment (e.g., a visible room), from the contributions resulting from seeing the possible space occupied by the sound sources (e.g., a panel hiding sound sources placed in front of the participant). In the first case, participants could code sound position with respect to existing references; in the second case, participants can develop some sort of visual prior about the position of the sounds in the environment (Parise, Knorre, & Ernst, 2014). It is worth noting that in most studies exploring spatial hearing, the position of the sound

sources is either directly visible or can be easily inferred. For instance, when all sources are hidden behind a curtain, participants can nonetheless infer that sounds can originate from a restricted portion of the space (Pavani et al., 2017). In this case, although participants have no detailed prior about the spatial layout of the speakers, they have continuous visual priors about the hemispace (front or back), elevation and distance of the speaker array.

One way to disentangle between these indirect visual contributions to sound localization is to exploit virtual reality. In a recent study, Ahrens et al. (2019) asked participants to perform a sound localization task in different visual scenarios created using virtual reality technology. In some conditions, visual information about the room was entirely prevented. In other conditions, participants were allowed to see a virtual version of the real room, comprising or not the speakers around them (Ahrens, Lund, Marschall, & Dau, 2019). Using VR, these authors were able to control the effect of both having the structure of the overall auditory environment and knowing the spatial likelihood of the auditory targets. They found that the reference frame provided by the visual information of the room without loudspeaker was enough to decrease error both in the horizontal and vertical dimensions compared to a blindfolded condition. Moreover, vision of the speaker array provided a further benefit compared to receiving only visual information of the room. Along a similar line, Majdak et al. (2010) have manipulated the whole visual background during a sound localization task to study the impact of seeing a simple visual spatial frame (a grid) compared to a condition of total darkness (Majdak et al., 2010). The results showed that even a simple grid improved precision in the horizontal plane. Furthermore, the visible grid reduced quadrant errors in the vertical plane, particularly in the front hemispace. In this work, sounds were always delivered using HRTF making more difficult to study the distance dimension. Likewise, the visible grid lacked binocular cues and did not provide information about distance. Finally, participants were forced into a static head listening posture. While this posture is most common in sound localization studies, it is a constraint that may limit the benefit from the visible visual environment on sounds localization. As proposed by some authors (Shelton & Searle, 1980), the spatial hearing facilitation that can result from seeing the overall environment, may reflect the active sensory-motor exploration of the auditory scene.

The present study aimed to test the contribution of a visual spatial frame to sound localization when participants are free to move their head without restriction. To this aim, we tested participants in two visual conditions: a uniform grey scene, in which no cues about the sound sources or the auditory environment were provided, and a simple visual spatial frame, in which a visible grid was the only visual information available about the environment. To present the different visual scenarios and to allow recording of spontaneous head and gaze movements, we took advantage of a new approach for the study of spatial hearing developed in our laboratories (Gaveau et al., 2020; Verdelet et al., 2020). Our approach is based on real sounds, virtual reality and real-time kinematic tracking and it allows: (1) accurate positioning of

real sounds at pre-determined locations with respect to the head; (2) measuring the participant's hand responses in the three dimensions of space (i.e., in 3D); (3) control over the visual scenarios and (4) free and measurable head-movements and gaze during sounds playback. The latter aspect is particularly advantageous for the present study, as we hypothesized that head and eyes orienting to the sounds could provide implicit measures of performance to further investigate the effect of visual manipulation. A secondary purpose of the present work was also to validate the feasibility of our VR approach for the study of sound localization abilities in humans.

2.3. Methods

2.3.1. Participants

Thirty-six participants (age: $M = 25.08$, $SD = 2.96$, range [20-32], 13 males, 34 right-handed) were recruited to participate in the experiment at the University of Trento (Italy), at the Integrative, Multisensory, Perception, Action and Cognition Team (IMPACT) lab in Lyon and at the center de Recherche Cerveau et Cognition (CerCo) of Toulouse (France). All participants signed an informed consent before starting the experiment, which was conducted according to the criteria of the Declaration of Helsinki (1964, amended in 2013) and approved by the respective ethical committees. All had normal or corrected-to-normal vision and reported no movement deficit. Hearing thresholds were measured using an audiometer for all participants, testing different frequencies (250, 500, 1000, 2000, 4000 Hz), on the right and left ear separately. All participants had an average threshold below 20 dB HL.

2.3.2. Apparatus and stimuli

Virtual reality (VR) and kinematic tracking was implemented using 3 identical HTC Vive Systems, one for each testing site. This method (European patent n°17723294.6–1115) has been developed in our laboratory (see also Gaveau et al., 2020; Valzolgher, Verdelet, et al., 2020) comprised one head-mounted display (HMD, resolution: 1080×1200 px, Field Of View (FOV): 110° , Refresh rate: 90 Hz), 1 controller (used by experimenter to calibrate head-position and to interrupt trial data registration), 2 trackers (one mounted on a short rod and held by participants to indicate the position of the sound and the other mounted above the speaker to track its position in real time) and 2 lighthouse base stations (scanning the position of the controller and trackers). Tracking precision and accuracy of the HTC Vive System is adequate for behavioral research purposes (Ahrens et al., 2019). The HMD was equipped with an SMI eye-tracking system (250 Hz). At all testing sites, stimuli were controlled and delivered using a LDLC ZALMAN PC (OS: Windows 10 (10.0.0) 64bit; Graphic card: NVIDIA GeForce GTX 1060 6GB; Processor: Intel Core i7–7700 K, Quad-Core 4.2 GHz/4.5 GHz Turbo - Cache 8 Mo - TDP 95 W) using Steam VR software and the development platform Unity3D (Unity Technologies, San Francisco, CA) Figure 66.

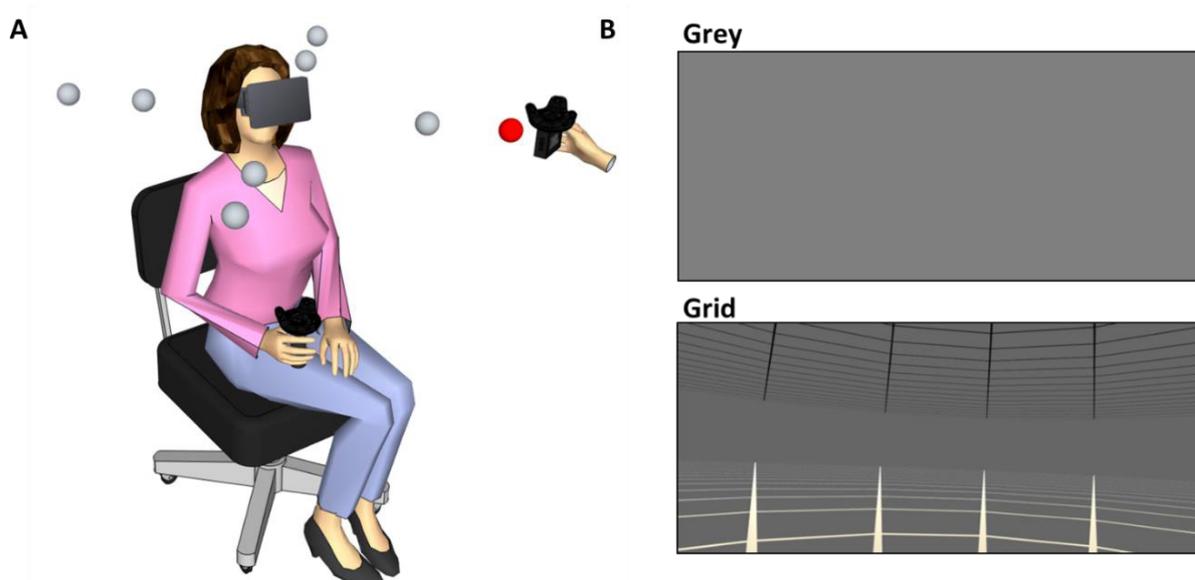


Figure 6 STUDY 1 - EXPERIMENTAL PROCEDURE AND SETTING

(A) Schematic representation of participant wearing the HMD and holding the tracker used for pointing responses; the eight spheres around the participant's head indicate the pre-determined speaker positions; the experimenter brought the tracked speaker (also shown in figure) at the pre-determined location identified in each trial. (B) Representation of the two 3D visual environments used in the study: grey and grid. A video showing dynamically the grid condition as seen from the participant's perspective is available at: <https://youtu.be/89xPLzr3fyQ>.

Participants were seated on a rotating armless chair with no chin rest, in the center of the room. The rooms had the following dimensions: Trento: 4 × 3 m; Lyon: 3 × 6 m; Toulouse: 3 × 5 m. All rooms were quiet, but none was specifically treated for being anechoic and sound-proof.

Real free field auditory stimuli were delivered by an unseen loudspeaker (JBL GO Portable, 68.3 × 82.7 × 30.8 mm, Output Power 3.0 W, Frequency response 180 Hz – 20 kHz, Signal-to-noise ration > 80 dB), whose position was continuously tracked in space. They were 3 s white noise bursts, amplitude-modulated at 4 Hz, and delivered at about 60 dB SPL, as measured from the participant's head (using Decibel Meter TES1350A placed at ears level).

This solution allowed us to track the position of the speaker, the hand of participant and the Head Mounted Display, via a dead reckoning process using gyroscope and accelerometer, plus a correction signal from the lighthouse system every 8.333 ms. Both tracking method allowed us to track this position with a frequency sample of 250 Hz. The software is designed to guide the experimenter to align the real loudspeaker (the sound source) with each of the 8 pre-determined position in the virtual environment in each trial.

2.3.3. Procedure

Before starting the experiment, participants were instructed about the task and informed about the use of the VR equipment. Specifically, participants were asked to listen carefully to each sound until it finished, and then to indicate its location in space using the tracker held in their right hand. They were informed

that sounds could be delivered anywhere in the 3D around them, always within a reaching space. They were also informed that, during sound emission, they were not allowed to move their hand (which rested on their laps) nor their trunk (which was oriented straight ahead). After sound emission, they could freely rotate the chair to indicate the sound source. Note that head-movements remain unconstrained both during and after sound emission, allowing the possibility of spontaneous active listening behavior (e.g., orienting the head in the direction of the sound).

Participants performed sound localization under two visual conditions: a uniform grey scene (grey) and a more structured scene with spatial references (grid) (Figure 6B and Video). The grid comprised two horizontally laid figures, drawn like spiderwebs of 50-meters radius, with 19 straight sides (angle around 20°) and 20 sub-figures plan separated by 2.5 m. The first horizontal web was placed at floor level ($Y = 0$ m) and the second was placed at 10 m height ($Y = 10$ m). In creating the grid, we aimed to obtain a structured environment, with distance clue that conveyed the idea of a vast and open space (hence, the 50-meter radius of the spiderwebs). In addition, we aimed to avoid any vertical line that could be used by participants as visual anchor during sound perception and/or during the pointing response. The overall grids were centered on the participant's position. A video showing dynamically the grid condition as seen from the participant's perspective is available at: <https://youtu.be/89xPLzr3fyQ> (note that the video does not represent an experimental trial, but only depicts the environment visible to the participants).

After the participant sat on the chair in the experimental room and wore the HMD, the experiment began the head-center calibration which was performed by collecting the 3D position of the two ears using the controller. These head-center coordinates served as origin of the polar frame of reference that defined speaker, head and gaze positions throughout the experiment. Then, the eye-tracker calibration was performed: participants were asked to follow a moving dot with their eyes. Both head and eye calibrations were repeated whenever the HMD was temporarily removed during the experiment (i.e., during pauses).

The loudspeaker position in 3D space was calculated for each trial, with reference to the center of the head. In this way, despite participants sat without any chin-rest, we could carefully control the position of each sound source with respect to the ears. Eight pre-determined positions were used throughout the experiment, resulting from the combination of 4 different azimuths (-45° , 45° , -135° or 135°), 2 different distances (35 cm or 55 cm) and a single elevation (0° , i.e., ear-level) (Figure 6).

In each experimental trial, the experimenter moved the loudspeaker to the desired position in 3D space, following visual instruction generated in real-time by the computer. Instructions conveyed the pre-determined azimuth and distance position for the speaker. These instructions were visible only to the experimenter, and they were delivered using the stimulus visualization monitor (IIYAMA PROLITE E2280HS 22", resolution: 1980×1080 , format 16:9) placed in the testing room. The monitor provided a

bird-eye view of the experimental room and conveyed the pre-determined position of the sound source for that trial (as a red circle) and the actual real-time position of the speaker (as tracker picture). Using this visual instruction, the experimenter reached for computer-determined position rapidly, keeping the speaker approximately at ear-level. The computer considered the loudspeaker correctly positioned when it entered a sphere of 6 cm diameter centered on the pre-determined sound position.

The noises produced by the experimenter while placing the speaker were minimal. Nonetheless, participants were explicitly informed to pay attention to the target sounds, as any other sound in the room could be deceiving. To prove this point at the beginning of the testing session, the experimenter showed how she could stay to the right of participants while delivering the sound from their left by stretching the arm and hand holding the speaker. This was a demonstration that occasional noise cues about the movement of the experimenter could provide misleading information and thus discourage participants to rely upon this information. Most importantly, pilot work in our laboratory showed that participants ($N=6$) cannot reliably point to the speaker when the same placement procedure is used but no target sound is delivered. In this scenario the three-dimensional vector distance between the speaker and the response is on average 73.3 cm (SD = 31.7 cm).

The computer delivered the target sound only when three concurrent criteria were met: (1) the loudspeaker was in the 3D position pre-determined for the trial; (2) the participant's head was facing straight ahead; (3) the participant's eyes were directed straight ahead. Participants complied with criterion 2 (head pointing straight ahead) and criterion 3 (eyes gazing straight ahead) by taking advantage of visual stimuli displayed in the HMD. At the beginning of each trial two crosses were presented to the participant: a white cross in the background indicated the desired position of the head and eyes; and a thin white cross, indicating the actual head-position of the participant. Participants were instructed to move their head to align the two crosses. When the alignment was achieved the thin cross turned blue. Likewise, participants were instructed to stare at the cross center, a feedback of their gaze location was given by a blue circle. Once the three criteria were achieved simultaneously (speaker position; head position and eye gaze), all visual feedback disappeared, and the sound was delivered. Participants were instructed to respond only after the end of the sound, bringing the tracker to the perceived location of the sound and holding it still a few seconds. The experimenter terminated the registration of the tracking by pressing a button on the controller. No feedback on performance was provided (similar procedure was used also in Gaveau's study (Gaveau et al., 2020), see also (Valzolgher, Verdelet, et al., 2020b).

The experimental session was organized in 4 successive blocks, with a pause between each block. Visual conditions (grey or grid) alternated between blocks of trials. Half of the participants followed a grid-grey-grey-grid sequence, whereas the other half followed a grey-grid-grid-grey sequence. Each block comprised 40 trials (i.e., 5 trials for each of the 8 pre-determined positions), resulting in a total of 160 trials

(i.e., 10 trials for each pre-determined position in each visual condition). The entire experimental protocol lasted approximately 45 min.

2.3.4. Analyses

The position of all tracked elements (loudspeaker, head center and direction, hand and eyes) was inspected manually for each trial. Loudspeaker position was calculated as the mean of x, y, z coordinates from the beginning of the sound to the end. Head and hand positions were analyzed using custom-made software for the kinematic analysis of movements, running on MATLAB R2019b.

To study head and gaze movements, we calculated the tangential velocity on the x, y, z axis (expressed in degrees of rotation) using two-points central difference derivative algorithm (Bahill & McDonald, 1983) with 5 points for the half-window. To determine the sequence of head and hand movements, the beginning and the end of all movements were automatically detected using a velocity threshold procedure ($10^\circ/\text{s}$ for head and 400 mm/s for eyes). The results of this procedure were inspected off-line and corrected manually, if necessary. This procedure served to establish the spatio-temporal profile of head and hand behavior and extract relevant parameters for subsequent analyses (reaction times (RT) of the first head movement and first eyes movement). Head movements below 10° were discarded. Importantly, it also served to reject all trials in which participants did not comply with the instructions (i.e., they made anticipatory hand movement during sound playback) or because of artefacts or lack of data.

All statistical analyses and data visualizations were performed using R, R-studio environment and JASP 0.9.1.0.

2.4. Results

Our approach to sound localization allowed to decompose localization errors in the three dimensions of space: azimuth, elevation and distance. We started by studying participants' ability to discriminate sound source location in the uniform grey condition – which we considered as baseline – and then we focused on our key experimental question, concerning the effects of seeing a visual grid. This two-step approach was motivated also by our interest to validate our VR approach to spatial hearing which, for several aspects, is novel with respect to other previous methods. Unlike classic works using real sounds, our approach allows free but carefully controlled positioning of the speaker in head-centered coordinates. Moreover, unlike classic auditory virtual reality studies, it uses actual sounds delivered in space and aligned with the visual reality scenario.

2.4.1. Hand pointing responses: the uniform grey condition

Sound localization responses for each participant are shown in Figure 7 A and B in bird-eye and lateral views, respectively, as a function of sound position. Participants clearly discriminated stimulus side (overall left/right discrimination errors = 1.0%, SD = 0.7), and made very few front/back confusions (overall errors = 1.4%, SD = 0.6). In addition, as shown in Figure 7 C, responses in distance were clearly segregated for near and far targets (near: M = 42.7 cm, SD = 10.8 cm; far: M = 58.6 cm, SD = 11.7 cm; $t(35) = 16.15, p < 0.001$).

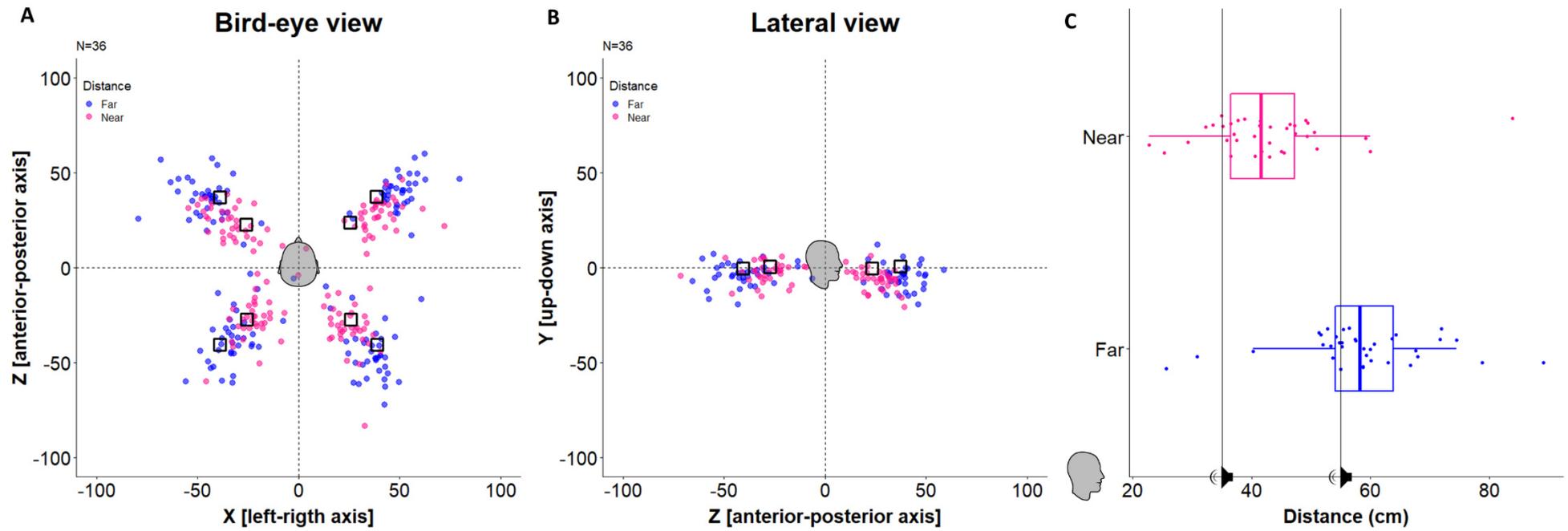


Figure 7 STUDY 1 - SOUND LOCALIZATION PERFORMANCE

(A) Bird-eye view of all target positions (squares with black border) and hand-pointing responses (smaller circles) in all trials and participants. (B) Lateral view of all target positions and responses. Responses for each participant are averaged across side (left or right) and distance (near, 35 cm and far, 55 cm). Responses are color-coded as a function of target distance (far is blue and near is pink). (C) Distance of participants' hand-pointing as a function of target position (Near or Far). Vertical lines represent the real position of the targets (Near: 35 cm; Far = 55 cm).

For each dimension, we studied absolute and signed errors as a function of target position, using an Analysis of Variance (ANOVA) with side (left or right), antero-posterior sector (front or back) and distance (near, 35 cm or far, 55 cm) as within-participants factors.

2.4.1.1. Azimuth

The overall absolute error in azimuth was 11.5° ($SD = 6.5^\circ$). The analysis on absolute error in azimuth revealed a main effect of antero-posterior sector, $F(1,35) = 6.00$, $p = 0.02$, $\eta^2 = 0.07$, caused by larger errors in back space ($M = 13.5^\circ$, $SD = 9.1^\circ$) compared to front space ($M = 9.4^\circ$, $SD = 7.2^\circ$). The main effect of distance also reached significance, $F(1,35) = 13.10$, $p < 0.001$, $\eta^2 = 0.02$. Errors were larger for near (35 cm) ($M = 12.6^\circ$, $SD = 7.8^\circ$) compared to far (55 cm) targets ($M = 10.4^\circ$, $SD = 5.5^\circ$). No other main effect or interaction reached significance (all $F_s < 2.51$, all $p_s > 0.12$).

A similar analysis on signed error (speaker position minus participant's tracker position, positive values indicate a rightward bias) revealed only a main effect of side, $F(1,35) = 26.39$, $p < 0.001$, $\eta^2 = 0.22$. This main effect reflects the fact that participants' pointing responses were more eccentric than actual sound azimuth (left: $M = -5.2^\circ$, $DS = 8.2^\circ$; right: $M = 5.3^\circ$, $DS = 6.4^\circ$) as shown in Fig. 2A. Thus, a positive (rightward) bias emerged for right sounds, where a negative (leftward) bias emerged for left sounds.

2.4.1.2. Elevation

The overall absolute error in elevation was 13.5° ($SD = 5.4^\circ$). The analysis on absolute error in elevation revealed a main effect of antero-posterior sector, $F(1,35) = 4.17$, $p = 0.05$, $\eta^2 = 0.04$, caused by larger errors in front space ($M = 14.9^\circ$, $SD = 6.8^\circ$) compared to back space ($M = 12.1^\circ$, $SD = 6.7^\circ$). In addition, there was a main effect of distance, $F(1,35) = 22.49$, $p < 0.001$, $\eta^2 = 0.05$. Elevation errors were larger for near ($M = 15.0^\circ$, $SD = 5.6^\circ$) compared to far sounds ($M = 12.0^\circ$, $SD = 5.8^\circ$). No other main effect or interaction reached significance (all $F_s < 3.27$, all $p_s > 0.08$).

A similar analysis on signed errors (positive values indicate an upward bias) revealed a main effect of antero-posterior sector, $F(1,35) = 25.15$, $p < 0.01$, $\eta^2 = 0.31$. Participants positioned the hand tracker below actual speaker location in front space (i.e., a downward bias; $M = -9.2$, $SD = 10.5$), but pointed above actual speaker location in back space (i.e., upward bias; $M = 4.2$, $SD = 8.7$). These biases were more pronounced for near than far targets, resulting in an interaction between antero-posterior sector and distance, $F(1,35) = 9.37$, $p = 0.004$, $\eta^2 = 0.01$.

2.4.1.3. Distance

The overall absolute error in distance was 12.8 cm ($SD = 6.7$ cm). The analysis on absolute error in distance revealed a main effect of side, $F(1,35) = 4.63$, $p < 0.04$, $\eta^2 = 0.02$, caused by larger errors for right ($M = 13.6$ cm, $SD = 7.5$ cm) than left targets ($M = 12.1$ cm, $SD = 6.6$ cm). In addition, all 2-way

interactions reached significance: side and distance, $F(1,35) = 13.43$, $p < 0.001$, $\eta^2 = 0.01$; side and antero-posterior sector, $F(1,35) = 8.06$, $p < 0.007$, $\eta^2 = 0.01$; antero-posterior sector and distance, $F(1,35) = 15.69$, $p < 0.001$, $\eta^2 = 0.02$. Overall these interactions capture the larger error in distance estimation that occurred for near targets in front right space ($M = 16.6$, $SD = 10.2$).

A similar analysis on signed errors in distance (positive values indicate overestimation of target distance) revealed the main effects of side (left or right), antero-posterior sector and distance all ($F_s > 9.87$, all $p_s < 0.003$). Overall, sound distance was overestimated ($M = 5.8$ cm, $SD = 10.7$ cm), but more for right ($M = 9.1$ cm, $SD = 10.5$ cm) compared to left sounds ($M = 2.5$ cm, $SD = 11.5$ cm). Overestimation was also larger for front ($M = 11.0$ cm, $SD = 10.7$ cm) compared to back sounds ($M = 0.6$ cm, $SD = 12.7$ cm), and for near ($M = 7.2$ cm, $SD = 10.8$ cm) compared to far sounds ($M = 4.4$ cm, $SD = 11.3$ cm). Several higher-order interactions also reached significance: side and antero-posterior sector, $F(1,35) = 5.65$, $p < 0.02$, $\eta^2 = 0.003$, side and distance, $F(1,35) = 5.46$, $p < 0.03$, $\eta^2 = 0.002$, and between side, antero-posterior sector and distance, $F(1,35) = 22.57$, $p < 0.001$, $\eta^2 = 0.005$. This 3-way interaction did not reveal any difference between near and far signed errors targets when placed in the back-right sector.²

2.4.2. Hand pointing responses: the effects of adding a visual frame

Having established performance in the three dimensions in the grey visual condition, we turned to examine to what extent adding a visual frame affected sound localization performance. To this aim, we entered absolute and signed errors into ANOVAs with visual condition (grey, grid) and antero-posterior sector as within-participant variables, separately for each spatial dimension. We focused on the antero-posterior position of sounds because we predicted that the effect of the visual manipulation could be maximal for sounds were delivered in the front space rather than back space.

The analyses on absolute and signed errors in azimuth, elevation and distance, did not reveal any main effect or interaction involving visual condition (all $F_s < 1.97$, all $p_s > 0.17$). One notable exception was signed error in elevation. For this measure, we found a significant interaction between antero-posterior sector and visual condition, $F(35) = 12.72$, $p = 0.001$, $\eta^2 = 0.27$. When sounds were delivered in the front space, participants judged their position lower than their actual location. This bias was smaller in the grid condition ($M = -6.8^\circ$, $DS = 10.3^\circ$) compared to the grey condition ($M = -9.2^\circ$; $DS = 10.5^\circ$; simple main effect:

² We interpreted these differences observed for near sounds in right space as the consequence of bio-mechanical constraints related to the fact that all participants were asked to point to the sounds using their right hand. To assess this hypothesis empirically, we analysed the data of a pilot experiment in which 12 participants performed the same task reported in the present manuscript but using their preferred hand on a trial-by-trial basis. An ANOVA on absolute errors in distance on this dataset, using again side, antero-posterior sector and distance as factors, yielded no significant main effect or interaction (all $F_s < 3.20$, all $p_s > 0.10$). On signed errors in distance, we found only a significant interaction between antero-posterior sector and distance, $F(1,11) = 7.22$, $p = 0.02$, $\eta^2 = 0.03$. This provides initial evidence that at least part of the specificities observed for right-sided sounds could depend upon the imposed use of the right hand for the response.

$F = 12.12, p = 0.001$). When sounds were delivered in the back space, participants judged their position as higher than their actual location. Again, this bias smaller in the grid condition ($M=2.7^\circ, DS = 8.6^\circ$) compared to the grey condition ($M=4.2^\circ; DS = 8.7^\circ$; simple main effect: $F = 5.84, p = 0.02$) (Figure 8 and Table 2).

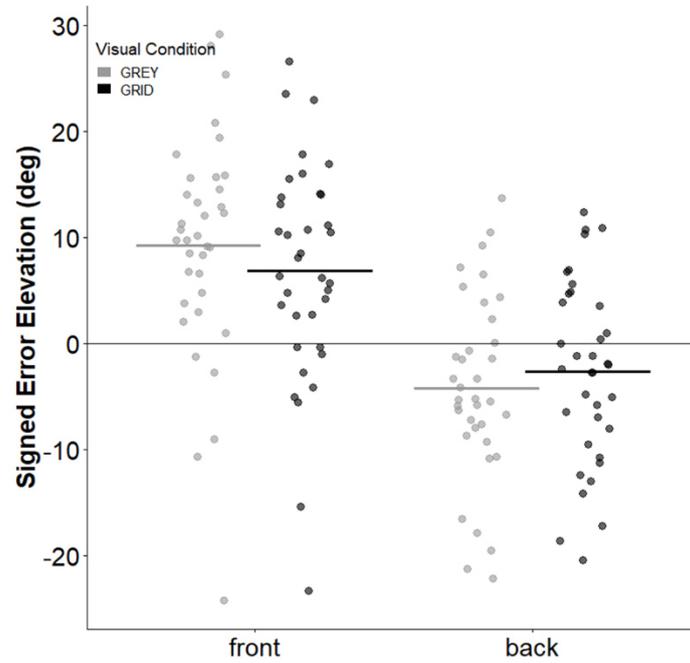


Figure 8 STUDY 1 - SIGNED ERROR IN ELEVATION

*Signed Error in Elevation (deg) as a function of Visual condition (grey, grid) and antero-posterior sector (front, back). In both graphs horizontal bars represent the mean of each condition, while points show participants value. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.*

		Azimuth (deg)		Elevation (deg)		Distance (cm)	
		Absolute Error	Signed Error	Absolute Error	Signed Error	Absolute Error	Signed Error
GREY							
NEAR							
FRONT							
	Left	10.3 (7.8)	-5.8 (9.2)	16.0 (8.1)	-12.0 (10.9)	12.8 (9.0)	9.8 (12.1)
	Right	11.5 (13.1)	4.9 (13.5)	17.7 (9.5)	-10.0 (15.4)	16.6 (10.2)	15.8 (10.9)
BACK							
	Left	15.8 (13.2)	-5.7 (16.0)	14.0 (8.8)	5.1 (12.3)	10.4 (7.6)	-1.0 (12.1)
	Right	12.8 (9.6)	6.6 (11.1)	12.2 (7.6)	4.7 (10.5)	11.8 (9.3)	4.4 (13.3)
FAR							
FRONT							
	Left	8.5 (6.1)	-4.5 (7.5)	12.7 (6.5)	-8.0 (10.7)	11.9 (9.3)	6.7 (12.9)
	Right	7.5 (6.2)	4.2 (5.3)	13.2 (8.1)	-6.7 (12.4)	14.0 (8.3)	11.8 (10.2)
BACK							
	Left	13.7 (12.9)	-4.7 (11.4)	11.5 (10.1)	3.2 (7.4)	13.2 (8.8)	-5.2 (14.6)
	Right	11.7 (7.2)	5.6 (10.1)	10.6 (7.2)	3.8 (10.3)	12.1 (8.5)	4.2 (13.5)
GRID							
NEAR							
FRONT							
	Left	11.2 (8.8)	-4.9 (9.4)	15.4 (8.3)	-8.7 (12.6)	12.9 (8.1)	10.8 (10.6)
	Right	9.3 (7.0)	6.4 (6.5)	15.7 (8.9)	-8.8 (14.0)	17.5 (10.1)	16.3 (11.7)
BACK							
	Left	14.8 (11.2)	-9.0 (14.2)	13.8 (7.9)	1.7 (12.8)	10.3 (7.7)	-1.2 (12.4)
	Right	12.4 (7.6)	7.3 (10.2)	12.2 (7.5)	3.8 (10.7)	12.6 (8.1)	5.1 (12.2)
FAR							
FRONT							
	Left	8.9 (6.4)	-4.9 (8.7)	11.8 (6.4)	-5.3 (10.1)	11.4 (8.6)	5.7 (12.8)
	Right	7.4 (8.8)	3.7 (6.0)	12.6 (8.8)	-4.6 (11.2)	14.3 (9.1)	12.0 (11.7)
BACK							
	Left	13.6 (9.5)	-6.3 (12.1)	11.6 (10.7)	2.6 (12.0)	13.9 (9.1)	-6.6 (12.9)
	Right	11.2 (5.6)	7.7 (7.4)	9.8 (7.3)	2.5 (8.8)	12.5 (10.3)	3.7 (15.3)

Table 2 STUDY 1 - ABSOLUTE ERROR AND SIGNED ERROR

Absolute error and Signed error of 3 dimensions of the space as a function of target position for grid and grey visual conditions. Standard deviation (SD) between brackets.

2.4.3. Head rotation

Head-movements occurred in 68.3% of trials on average (SD = 38.3%; median = 89.9, range = 0–100%), for targets in front and back space (71.3% and 65.3% of trials, respectively). Figure 9A shows the substantial variability of this spontaneous behavior in both visual conditions. In front space, the first head movement for sounds at 45° to the right was of 40.1° (SD = 7.2°), whereas for sounds at 45° to the left it was 39.2° (SD = 7.3°). In back space, the first head rotation for sounds at 135° to the right was of 67.9° degrees (SD = 21.0°), whereas for sounds at 135° to the left it was 68.7° (SD = 22.5°) (see Figure 9 B).

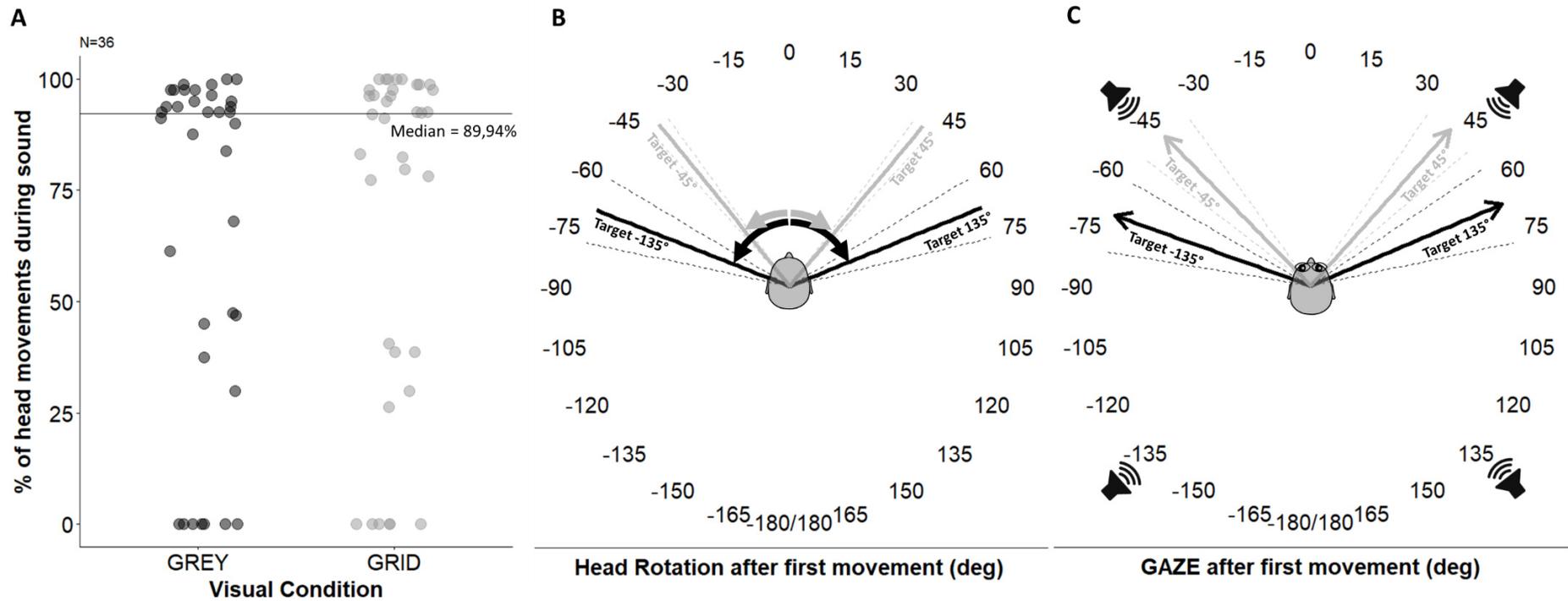


Figure 9 STUDY 1 - HEAD ROTATION

(A) Boxplot of percentage head-movements: the percentage of trials in which head movements occurred for each participant. (B) Average of the rotation around vertical axis of the first head movement only when participants turned to the side of the target (same dataset of the analysis). (C) Average of the direction of the first gaze movement only when participants turned to the side of the target (same dataset of the analysis). Dashed lines represent confidence intervals.

To analyze head movements rotation around the vertical axis, we considered two variables: 1) rotation amplitude of the first head movement (deg) and 2) time to first head-movement (the time between the beginning of the sound emission and the beginning of the first head movement in millisecond). Trial without head movements were excluded from these analyses. This resulted in the exclusion of 6 participants which did not move at all and 5 participants for whom less than 12 trials per condition remained. In addition, we focused only on those trials in which participants rotated towards the correct sound side. This occurred on 97.5% (SD = 2.0) of trials on average, indicating again that participants easily disambiguate sound side.

To test the effect of the presence of the grid on these two variables (head-rotation latency and amplitude), we entered each measure separately in an ANOVA with visual condition and antero-posterior sector as within variables. ANOVA on head-rotation amplitude revealed no main effect or interaction (all $F_s < 1.26$, all $p_s > 0.27$). The analysis on time to first head-movement revealed that participants started to rotate their head 978 milliseconds (SD = 155) after sound emission on average. In addition, a significant main effect of antero-posterior sector emerged, $F(1,24) = 7.90$, $p = 0.02$, $\eta^2 = 0.25$. When sounds were emitted in front space, participants responded faster (M= 957 ms, SD = 154, ms) compared to when they were emitted from the back (M= 999 ms, SD = 156 ms) (Figure 10).

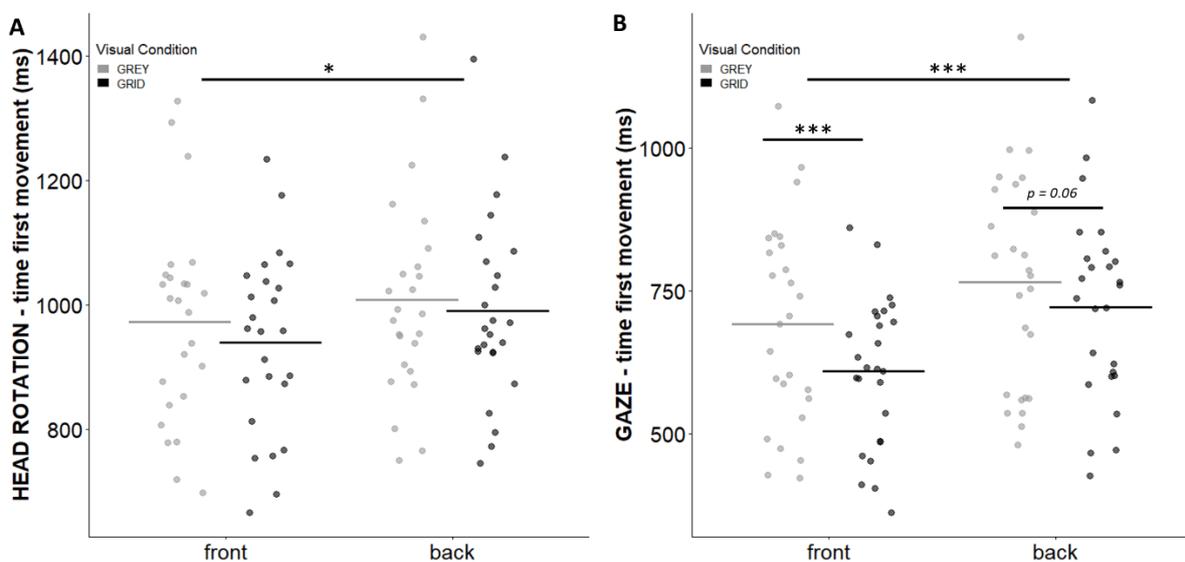


Figure 10 STUDY 1 - HEAD AND GAZE FIRST MOVEMENTS

The time between the beginning of the sound stimulation and the beginning of the first head or eyes movement. (A) Boxplot of Head Rotation time first movement as a function of target position (front, back) color coded as a function of Visual Condition (grey, grid). (B) Boxplot of GAZE time first movement as a function of target position (front, back) color coded as a function of Visual Condition (grey, grid).

2.4.4. Gaze direction

Similar to head rotation analyses, we focused on trials in which participants gazed towards the same side of the sound (correct trials). This resulted in the exclusion of 10 participants for whom less than 12 trials per condition remained. On average participants gazed towards the same side of the target on 97.2% (SD = 1.7) of trials. Figure 10 B shows gaze direction as a function of target position. In front space, first gaze was directed to 42.1° (SD = 18.9°) for sounds 45° right, and to 44.6° (SD = 20.7°) for sounds 45° left. In back space, first gaze was directed 67.2° (SD = 27.3°) for sounds 135° right, and to 70.5° (SD = 25.2°) for sounds 135° left.

To analyze gaze movements, we considered two variables: (1) direction of the first gaze movement (deg) and (2) time to first gaze-movement (the time between the beginning of the sound emission and the beginning of the first gaze movement in millisecond).

To test the effect of our visual manipulation we entered these two dependent variables in separate ANOVAs, with visual condition and antero-posterior sector as within variables. The ANOVA on direction of the first gaze movement revealed a significant two-way interaction, $F(1,25) = 5.70$, $p = 0.03$, $\eta^2 = 0.19$, caused by a less symmetrical gaze directions in response to sounds in back space ($M = -3.4^\circ$, $DS = 9.0^\circ$), specifically in the grid condition (simple main effect: $F = 5.01$, $p = 0.03$).

A similar ANOVA on time to first gaze-movement revealed that participants started to gaze sounds 698 milliseconds (SD = 164) after sound emission. In addition, we found a significant main effect of antero-posterior sector, $F(1,25) = 19.54$, $p < 0.001$, $\eta^2 = 0.44$. Unsurprisingly, participants oriented their gaze faster to front ($M = 651$ ms, $SD = 153$ ms) compared back sounds ($M = 744$ ms, $SD = 174$ ms) (Figure 10 B). More interestingly, we also found a significant main effect of visual condition, $F(1,25) = 12.85$, $p = 0.001$, $\eta^2 = 0.34$. When the visual scenario was the grid, participants latency was shorter ($M = 666$ ms, $SD = 146$ ms) compared to when it was grey ($M = 728$ ms, $SD = 271$ ms). To test specifically whether visual condition had a greater impact in front than back space we run simple main effects. These revealed that the effect of visual condition was more substantial in the front ($F = 19.46$, $p < 0.001$) than back space ($F = 3.82$, $p = 0.06$; note that the 2-way interaction between visual condition and antero-posterior sector also approached significance: $F(1,25) = 6.46$, $p = 0.075$, $\eta^2 = 0.12$) (Figure 10 B).

2.5. Discussion

The present study examined the effect of seeing a simple visual spatial frame on sound localization, in a context in which participants were free to move their head and eyes while listening to sounds. To this aim, we leveraged a virtual reality (VR) approach which allows accurate control over the visual scenarios,

accurate playback of free field sounds in three-dimensional space, and free and measurable movements of head and gaze during sounds playback.

When no visual spatial frame was available (grey condition), sound localization in azimuth was better in front than back space. When placing the tracker in back space participants had to turn around and modify their posture. This adjustment entails re-coding of sound position from egocentric to world-centered coordinates, leading to more uncertain estimates of sound position (Kopinska & Harris, 2003; Pavani, Husain, & Driver, 2008). In addition, during the grey condition the participants' responses were characterized by a bias to point to positions more eccentric than actual speaker locations. Errors in azimuth were higher for far rather than near targets. Although elevation was not varied, errors in the vertical dimension were larger for near than far sounds, and frontal sounds were perceived as lower than their actual position. Sound distance was overestimated, particularly for front and right sounds. Concerning distance, it is important to notice that our approach led us to measure sounds localization abilities also considering this dimension.

Adding a visual spatial frame (grid condition) did not change the overall pattern of hand pointing responses in azimuth or distance. However, it resulted in improved accuracy in elevation and it affected participant's gaze behavior, producing faster orienting responses to the sounds. The question on the role played by visuo-spatial references on sound localization dates back to the 1970s (Warren, 1970), when it was termed 'visual facilitation'. Yet, until the advent of VR methods, visual facilitation could only be addressed with somewhat crude manipulations (e.g., blindfolding participants). Most importantly, when visual information about the environment was provided (i.e., eyes-open conditions), it was a methodological challenge to disentangle the contributions related to seeing the overall structure of the environment, from the contribution related to seeing the potential sources of sound. Using a VR approach, we investigated the role of visually structured information on sound localization, in the absence of visual priors about the speakers' position. While immersed in the VR scene, participants were only informed that target sounds would be delivered within reaching distance but had no further information on their positions — i.e., they expected sounds to appear all around the body but inside an estimated range of distance (Gaveau et al., 2020).

To the best of our knowledge, only two previous works have addressed a similar question using VR techniques (Ahrens et al., 2019; Majdak et al., 2010). Majdak et al. (2010) tested the effect of seeing a simple visual environment on sound localization. Subjects were immersed in a virtual environment which comprised a sphere (diameter 5 m), marked with grid lines every 5° horizontally and every 11.25° vertically. Furthermore, participants required to judge elevation and azimuth of the sound taking advantage of grid lines. They reported that participants' hand pointing to sounds was enhanced by the visual grid in both horizontal and vertical dimensions, compared to a condition of total darkness. Our

results are consistent with these findings. However, in the present study we observed an advantage only in terms of elevation, whereas sound localization in azimuth remained unchanged in the presence of the visual grid. This different result could reflect methodological discrepancies between the two studies. First, our visual grid was intentionally conceived to avoid any vertical line that could serve as anchor for sound localization. Second, our virtual environment conveyed also a sense of distance, which could have introduced additional uncertainty in the interpretation of the auditory cues. Third, we used real sounds instead of virtual ones, possibly leading to more precise sound localization overall.

Nonetheless, both studies converge in stressing the importance of having a visual frame when localizing sound. In particular, the grid condition may have provided a detailed visual map for positioning sound sources.

Ahrens and colleagues also took advantage of VR (Ahrens et al., 2019), allowing participants to see the room dimensions as well as their hand-position, whereas the speakers' array was not visible. Using sounds beyond reaching distance (2.4 m from the head), they found that visual information decreased errors both in the horizontal and vertical dimensions, as compared to a blindfolded condition. In Ahrens et al.'s study the VR reconstructed room was a careful visual replica of the in which the experiment was conducted, with the foam wedges of the anechoic chamber providing extremely rich vertical and horizontal visual references all around the participant. As discussed above for the study by Majdak and colleagues, it may be that these substantial visual cues played a role in the improvements observed in the horizontal dimension, serving as references or place-holders for sound localization when visible. In our more minimalist visual scenario we intentionally avoided all vertical visual references (Majdak et al., 2010).

One difference with respect to both these previous works is that we presented sounds within reaching distance. In Majdak et al., sounds were delivered through headphones and although they were likely perceived externalized, it is difficult to establish at which distance they were perceived. In Ahrens et al., sounds were instead delivered away from the body, at 2.4 m. Our rationale in presenting sounds within reaching distance was to allow participants to respond using the hand-held tracker and measure their accuracy in distance estimation. However, this choice limited our studied space to the near-field and potentially influenced the interactions with the visual environment we created. Future works could examine to what extent the impact of visual environment on sound localization could emerge differently for sounds at different distances from the body, or — in case of enclosed spaces like rooms — at different relative distances with respect to the visible surfaces.

Another important difference is that our participants were free to move their heads while listening to sounds. This spontaneous orienting behavior also involved the eyes, and likely made azimuth localization easier for our participants. In turn, this could have reduced the possibility of observing visual condition

effects in azimuth. However, this gaze (i.e., head and eyes) orienting behavior revealed interesting findings. We found that correct gaze orienting responses in the direction of the sound started earlier when exposed to a visual grid compared to the grey condition. A similar trend was observed also for head rotation. We believe that the interest of analyzing this measure is related to the fact that, unlike hand pointing, gaze orienting is a more implicit measure of sound localization performance — particularly when the first gaze to sound response is considered. Measuring participants' gaze allowed us to capture an early and implicit aspect of acoustic space perception, which differs from the explicit head pointing method that has been used in sound localization tasks as an alternative to hand pointing. Taking this perspective, our results suggest that even a minimal visual spatial frame can speed up the right-left implicit disambiguation of sound position.

These effects of the visual features of the environment on sound localization are complementary with the line of research that examined the effect of the *acoustic* feature of the environment on sound localization and visual scene perception. For instance, Gil-Carvajal et al. (2016) have examined the effect of a mismatch between playback and recording room on perceived distance and azimuth direction of sounds (Gil-Carvajal, Cubick, Santurette, & Dau, 2016). They found that sound distance ratings decreased when measured in an environment that was more reverberant than the original recording room, whereas azimuth direction remained unaffected. Most interestingly, they also observed that changes in the visual attributes of the room were ineffective and concluded that visual congruency is less crucial than the correspondence between the acoustical features of environment and the stimuli.

Other works examined the interplay between auditory and visual features of the environment. Etchemendy et al. (2017) have found that auditory reverberation cues impact on the perception of visible room size, revealing that the auditory environmental context can modulate visual distance perception (Etchemendy et al., 2017). At the same time, Schutte et al. (2019) have shown that visual room impression does not affect people's abilities to estimate rooms' reverberations (Schutte, Ewert, & Wiegrebe, 2019). The latter evidence in particular is directly relevant to the present work, as it suggests that the effect of the visual grid proposed in our experiment may not have influenced the extraction of the acoustic features from the environment.

In the present study conducted in three different experimental sites, we did not measure the rooms' reverberation limiting the possibility to deepen the acoustic influences on both sound localization performance and visual manipulation effects. Further work should aim to combine the contribution of visual and auditory features of the environment with the active sound localization approach introduced by the present work, to examine to what extent the relative contribution and potential interactions between these multisensory contextual contributions to spatial hearing. It would also be important to assess sound localization beyond reaching (as here) to address more directly whether sounds that perceived further

away from the head (e.g., at 2.4 m as in Ahrens et al., 2019 or beyond) could interact more with the wider spaces of rooms or open VR environments like the ones simulated in our grid condition.

2.6. Conclusion

In this study, we documented that providing visual frame that is totally uninformative about sound source position in space helps sound source localization in active listening conditions. These findings contribute to emphasize the indirect but positive contribution of minimally structured vision to sound localization and further promote the idea that sound localization ability should be conceived as a multisensory process. Our findings also underlie the importance of allowing and measuring spontaneous head and gaze movements. The here adopted VR technology allowed us to go beyond traditional approaches in the study of spatial hearing by allowing participants to move their eyes and head during and after sound presentation, while retaining full control over sound placement and recording of our dependent variables using ecologically valid contexts, which is crucial in studying hearing experience as suggesting by recent studies in the field (Hadley, Brimijoin, & Whitmer, 2019).

Chapter 3. Study 2: Spontaneous head-movements improve sound localization in ageing adults with presbycusis

3.1. Abstract

The aim of this study was to examine if spontaneous head movements during sound emission can improve sound localization in ageing adults with presbycusis. Using motion-tracking in combination with real free-field sounds delivered in a visual virtual reality scenario, we tested hearing-impaired ageing adults in a front-back discrimination task and in 3D sound localization task in the frontal plane. Participants completed the tasks either with their head either static or free to move (active listening). We found that active listening improved front-back discrimination and reduced localization errors in the frontal plane, in azimuth and elevation but not distance perception. These findings extend to ageing adults with presbycusis the limited literature on the advantage of head-movements on sound localization, by providing initial evidence that the advantage of active listening persists in ageing and can prove effective even in case of mild hearing loss. These results could be exploited when planning interventions in this hearing-impaired population, that could exploit self-regulation strategies and active behavior.

3.2. Introduction

Sound localization is the ability to determine the location of a sound source. When perceiving acoustic space, our brain combines the auditory cues that result from the interactions of sounds with the head and external ears (Middlebrooks, 2015) and creates an association between these cues and external spatial coordinates (van der Heijden et al., 2019). This ability is greatly affected by age-related hearing loss (i.e., presbycusis). Presbycusis can change relative intensity and time of sounds at the two ears (binaural cues) and it can affect sound frequency perception at each single ear (monaural or spectral cues). Binaural cues play a key role for sound localization in the horizontal dimension. Monaural cues depend upon the direction of sounds with respect to the head and the pinnae, and contribute to front-back disambiguation, elevation estimation, and distance perception of sounds (Middlebrooks, 2015; Middlebrooks & Green, 1991; Risoud et al., 2018). Because high frequencies (>4 kHz) are the first to be affected by presbycusis (Rakerd et al., 1998) and greatly contribute to monaural cues, it is common for people with age related hearing loss to have poor performance in vertical localization (Noble et al., 1994) and in front-back discrimination (Middlebrooks, 2015; Middlebrooks & Green, 1991; Risoud et al., 2018). In addition, because high-frequencies contribute primarily to Interaural-Intensity Differences (IID), presbycusis reduces the available binaural cues for sound localization along the horizontal axis (Abel, Giguère, Consoli, & Papsin, 2000). Finally, asymmetric hearing loss, could disrupt binaural cues even further.

To date, the interventions proposed can only partially resolve this problem. For example, hearing aids can improve listening by amplifying the missing high frequencies and facilitate speech understanding, especially in silence. However, they do not fully contribute to better sound localization and spatial hearing

problems persist (Denk, Ewert, & Kollmeier, 2019). Altered spatial hearing impacts on two fundamental abilities: listening in noise (acoustic space contributes to perceive the acoustic scene) and visual attention orienting (sounds guide visual attention). The present study is pilot research aiming at evaluating a further perspective to improve sound localization abilities of people with presbycusis: the ability to move the head while listening.

An aspect that is not fully appreciated is that sound localization abilities are usually tested asking participants to keep their head still (Litovsky, Parkinson, & Arcaroli, 2009; Távora-Vieira, De Ceulaer, Govaerts, & Rajan, 2015; van Hoesel & Tyler, 2003). In many cases, the head is held in place by a chinrest (Brungart, Durlach, & Rabinowitz, 1999; Pavani et al., 2008). This head-fixed situation, however, hardly occur in everyday life. Hence, these conditions (1) may provide a partial measure of the individual's sound localization abilities, and (2) may limit our understanding of the potentials for improvement related to head movements. Although the fundamental role of head-movements in spatial hearing has been advocated since the first half of the last century (Wallach, 1940), it has become the object of more systematic investigations only in the last decades. When head-movements are allowed, young adults with normal hearing can improve their ability to localize sounds in the horizontal and vertical dimensions (Perrett & Noble, 1997a). In addition, compared to a condition in which listeners are required to keep their heads still, active head-movements during sound emission can reduce front-back confusions in sound localization (Iwaya et al., 2003). A similar advantage has been shown also in hearing-impaired people (Brimijoin, McShefferty & Akeroyd, 2012) and in people listening with cochlear implants (Pastore, Natale, Yost, & Dorman, 2018). In the context of presbycusis, this notion is relevant because head movements could help to compensate for the distorted auditory inputs, by providing richer auditory cues and by allowing active listening strategies – i.e., self-regulation behaviors aimed at collecting more information from the auditory environment. Yet, the role of head-movements in spatial hearing of individuals with presbycusis remains largely overlooked.

The present study aimed to examine if head movements could improve sound localization in ageing people with presbycusis as compared to a head-static listening condition. In addressing this question, we took advantage of a novel virtual reality and motion-tracking approach to sound localization (Coudert et al., 2021; Valzolgher, Alzhaler, et al., 2020; Valzolgher, Verdelet, et al., 2020). This methodology which allows extensive control over the visual and auditory stimulation delivered to the participant, while recording head-movements in real-time. We tested elderly participants in a sound localization task in the frontal plane and a front-back discrimination task. Across tasks, sound position changed horizontally, vertically and in distance in the frontal plane, or along the front-back axis. We measured performance in a within-experimental design under two conditions: one in which ageing participants were free to move their head during sounds delivery (active listening), and one in which they were asked to remain still during

sound delivery (static listening). Based on previous studies on young hearing participants (Iwaya et al., 2003; Perrett & Noble, 1997a) and young listeners with hearing deficits (Brimijoin & Akeroyd, 2012; Pastore et al., 2018), we expect that head movements should facilitate sound localization, both in the front-back discrimination task and in the sound localization task in the frontal plane.

3.3. Material and Methods

3.3.1. Participants

Eight participants (mean age 70.5, SD = 6.50, range = [62 - 81], 5 males) took part in the study. They were selected by the otolaryngology department of S. Maria del Carmine hospital in Rovereto (province of Trento, Italy), according to the following inclusion and exclusion criteria. They had to be native Italian speakers aged between sixty-five and eighty, with normal or correct-to-normal vision. Critically, they were selected for the presence of hearing loss: in particular, we focused on individuals with moderate (41 – 60 dB HL) to severe (61 – 80 dB HL) sensorineural hearing loss, who also presented an asymmetry of hearing loss between the two ears. The rationale for this choice was to recruit participants for whom both monaural and binaural auditory cues could be impacted by presbycusis. Participants were excluded if: they were younger or older than that declared in the inclusion criterion; they used hearing aids; they had a psychopharmacological therapy in progress; they had taken drugs with side effects on psychic and attentional functions; they were suffering from diseases that could interfere with the task's execution (neurological or psychiatric pathologies particularly); they were unable to provide informed consent and could not perform the tasks required by the experimental procedure.

Those who agreed to take part in the study were contacted by the researchers, who explained the planned tasks in detail. All received a diagnosis of asymmetrical hearing loss: for five of them it was moderate, for three severe. The average difference between hearing thresholds of the two ears across participants was 42.75 dB HL (SD = 11.21, range = [25 - 60]). The average hearing threshold in the worse ear was 60.88 dB HL (SD = 7.94, range = [51 – 78]), whereas the best ear was 21.50 dB HL (SD = 7.31, range = [15 - 38]). Their audiometry was provided to us by hospitals technicians. For five participants the worst ear was the left one, for the remaining three the worse ear was the right one. Further information is provided in the Table 3.

All volunteers gave their informed consent before starting the experiment, which was approved by the Ethic Committee of the University of Trento (protocol number: 2019-037).

Participant	Age	Sex	HL	Left (dB HL)	Right (dB HL)	Hearing thresholds difference (dB HL)
1	68	M	Severe	63	15	48
2	69	M	Moderate	58	21	37
3	77	M	Moderate	16	58	42
4	71	F	Moderate	23	51	55
5	81	F	Severe	63	38	25
6	62	M	Moderate	60	23	37
7	63	M	Severe	18	78	60
8	73	F	Moderate	56	18	38

Table 3 *STUDY 2 - ANAMNESTIC AND AUDITORY CHARACTERISTICS OF PARTICIPANTS*

3.3.2. Stimuli

Each auditory target was a 5 seconds white-noise burst, amplitude-modulated at 2.5 Hz to enhance auditory cues to sound localization. It was delivered at about 75 dB SPL, as measured from the participant's head using a Decibel Meter TES1350A placed at ears level. The 5 seconds duration was chosen to allow participants enough time to search the sound with head-movements in the active listening condition.

3.3.3. Apparatus

All stimuli were delivered within reaching space, using a single real speaker (JBL GO Portable, 68.3 × 82.7 × 30.8 mm, Output Power 3.0 W, Frequency response 180 Hz – 20 kHz, Signal-to-noise ratio >80 dB). The speaker was tracked in real-time within a calibrated space and moved manually by the experimenter to accurately reproducible target positions in 3-dimensional space.

The entire experiment was run using a PC that controlled a HTC Vive system, which is a virtual reality and motion tracking device, adapted for the study of spatial hearing by our laboratory (Coudert et al., 2021; Valzolgher, Alzhaher, et al., 2020; Valzolgher, Verdelet, et al., 2020a). The system comprised one Head-Mounted Display (HMD, resolution: 1080 x 1200 px, Field of View (FOV): 110°, Refresh rate: 90 Hz) used for presenting all visual scenarios and stimuli; one hand-held tracker used by participants to indicate the perceived sound's location; one tracker placed above the speaker to monitor its position in real-time; one hand-held controller used by the experimenter to record the responses; and two lighthouse base stations for scanning of the position of the trackers. Our approach allows to present real free field sounds at reproducible horizontal, vertical and depth positions in the space around the participant.

Stimuli were delivered by a LDLC ZALMAN PC (OS: Windows 10 (10.0.0) 64bit; Graphic card: NVIDIA GeForce GTX 1060 6 GB; Processor: Intel Core i7-770K, Quad-Core 4.2 GHz/4.5 GHz Turbo – Cache 8 Mo – TDP 95W) using Steam VR software and the development platform Unity. All target positions were defined with respect to the center of the head and the inter-aural axis and updated in real-time in case of changes in head position. For instance, a target position at 40 degrees to the left, 20 degrees below ear level and at 35 cm from the center of the head can be stimulated repeatedly across trials, even in case the participant's head changes in position. The procedure to guide the speaker to the target positions, with accurate and reproducible placements across trials has been documented elsewhere (Coudert et al., 2021; Valzolgher, Alzhaler, et al., 2020; Valzolgher, Verdelet, et al., 2020). This approach allows a natural posture for the participant, with unrestrained initial head position and without imposing head movements on a single plane (Perrett & Noble, 1997a).

3.3.4. Procedure

3.3.4.1. General procedure

The entire experiment was completed in a single session that lasted about 2 hours. Prior to the testing session, each participant signed the informed consent to participate in the study. Next, they completed a questionnaire aimed at collecting anamnestic data and were administered the Montreal Cognitive Assessment test to exclude a possible cognitive impairment (Conti, Bonazzi, Laiacona, Masina, & Coralli, 2015).

After these preliminary steps, participants were asked to sit on a rotating chair fixed in the center of the room (3 x 4 x 5 meters) and to wear the HMD. At this stage, as well as each time the HMD was removed and then put back on again, the position of the participant's head was calibrated, so that all target positions could remain precisely anchored to the center of the head and the inter-aural axis. This moment was followed by three phases: (1) the adaptation to the virtual visual environment and to posture requests; (2) the front-back discrimination task; (3) the sound localization task in the frontal plane. Note that the order of the two localization tasks was counterbalanced across participants.

3.3.4.2. Adaptation to the virtual environment and to posture requests.

Upon wearing the HMD participants found themselves immersed into a virtual room, with the exact same metrics as the real room (i.e., 3 x 4 x 5 meters) but completely empty. This choice was motivated by previous studies in which it was shown that spontaneous head movements occur more in the presence of a visual stimulation (Lederman & Klatzky, 2009; Redon & Hay, 2005). Moreover, the virtual room is a more ecological and less disturbing context than a dark environment with no visual references. In this virtual visual environment, the speaker and the experimenter were entirely invisible. Only the tracker that

the participants held in their hand was visible, to ensure that participants could accurately control it when pointing to the position of sounds. It should be noted that virtual reality has been used previously as experimental device when testing elderly participants, with no harmful outcomes or stressful situations reported (Crespo, Idrovo, Rodrigues, & Pereira, 2016; García-Betances, Jiménez-Mixco, Arredondo, & Cabrera-Umpiérrez, 2015; Molina, Ricci, De Moraes, & Perracini, 2014; Plancher, Gyselinck, Nicolas, & Piolino, 2010; Zonooz et al., 2019).

At the beginning of each task and block, 10 practice trials were included to allow participants to familiarize with the procedure. Whenever a new trial began, participants were required to align their heading direction (indicated by a white cross) with a white cross that was fixed in the center of the virtual room. The target sound was emitted only when the correct alignment was achieved. During the active listening condition, no suggestions were given regarding how to move the head and in this sense the movement represents a spontaneous strategy³.

3.3.4.3. Front-back discrimination task

In the front-back discrimination task, target positions were arranged along the participant's mid-sagittal plane, two at the front and two at the back (see Figure 11). The positions varied horizontally (at 0° or 180° from the subject's median line, corresponding to 0°), on the vertical plane (at 0° or 45° from participants' ear line, corresponding to 0°), and in distance (at 0.35 m or 0.55 m from the participant). A total of 64 targets were presented, equally distributed between the active and static listening conditions. Participants were instructed to wait for the sound to finish and then report if the sound source was in front of or behind them. In one block of trials, the participant was instructed to keep the head still (static listening); in the other block, the participant was allowed to move the head in the attempt to localize the sound (active listening). The experimenter pressed a key on the hand-held controller to record the response and trigger the next trial. No performance feedback was provided.

³ Participants were explicitly informed to pay attention to target sounds only, since any other sound in the room could be deceiving. The noises produced by the experimenter while placing the speaker were minimal. In fact, prior to the experiment, the experimenter showed how she could stay to the right of participants while delivering the sound from their left by stretching the arm and holding the speaker with her hand. Thus, she demonstrated that occasional noise cues generated by experimenter's movements could provide mis-leading information and therefore discouraged participants to rely upon this information. Most importantly, pilot work in our laboratory showed that participants (N = 6) cannot reliably point to the speaker when the same placement procedure is used but no target sound is delivered. In this scenario the three-dimensional vector distance between the speaker and the response is on average 73.3 cm (SD = 31.7 cm).

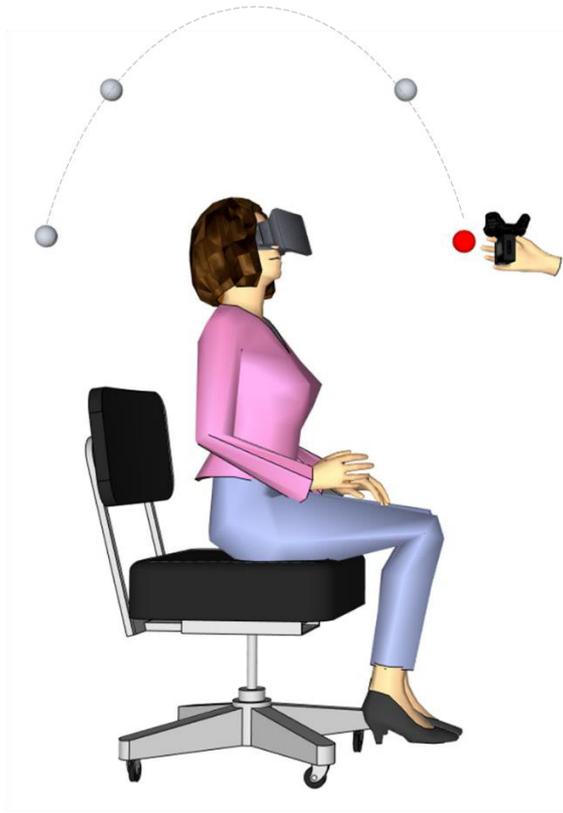


Figure 11 *STUDY 2 - FRONT-BACK DISCRIMINATION SETTING*
 Schematic description of participant wearing the hmd in the front-back discrimination task.
 The spheres above the participant's head indicate the target positions. The experimenter brought the tracked speaker (also shown in figure) at the pre-determined location identified in each trial.

3.3.4.4. Sound localization task in the frontal-plane

In the sound localization task in the frontal plane, target positions were all in front space, but at different azimuth, elevation, and depth (see Figure 12). Specifically, they were positioned: in the horizontal plane at $\pm 40^\circ$ and $\pm 20^\circ$ with respect to the midsagittal plane (corresponding to 0°); in the vertical plane at $\pm 20^\circ$ with respect to the plane passing through the ears (corresponding to 0°); in distance at 35 or 55 centimeters from the center of the subject's head.

Participants were instructed to listen to each sound and wait until its end before responding. While the sound was delivered, the participants kept their right hand holding the tracker stationary on the chest. At the end of the stimulation, they moved it to the perceived position of the sound. Participants were asked to stop in the selected position for a few seconds and then to press a button on the tracker to confirm their response. When this happened, the experimenter ended the trial and triggered the beginning of the next trial. No performance feedback was provided (Coudert et al., 2021; Valzolgher, Alzhaher, et al., 2020; Valzolgher, Verdelet, et al., 2020a).

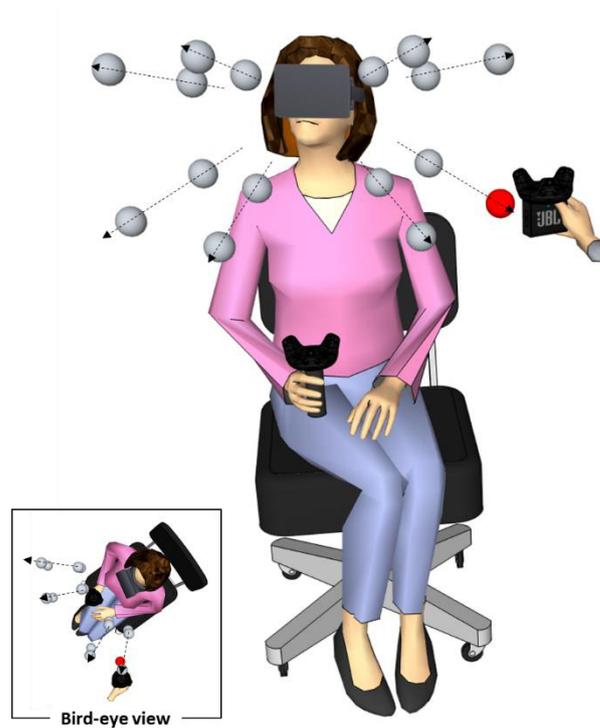


Figure 12 STUDY 2 - SOUND LOCALIZATION SETTING

Schematic description of participant wearing the hmd in the sound localization task.

The spheres around the participant's head indicate the pre-determined speaker positions. The experimenter brought the tracked speaker (also shown in figure) at the pre-determined location identified in each trial. (A) Frontal view; (B) Bird-eye view. Participant holds the tracker used for their pointing responses.

As for the front-back discrimination task, participants completed the task in two listening conditions that changed across blocks: with the head held fixed (static listening) or with the head free to move during sound presentation (active listening). Overall, the task comprised 96 trials (i.e., 3 repetitions for each combination of the 4 azimuths x 2 elevations x 2 distances), equally divided in the two listening condition blocks.

3.3.5. Analysis

In the front-back discrimination task, the dependent variable was the number of correct responses reported in the two listening conditions. In the sound localization task in the frontal plane, the dependent variables were the average 3D error (in centimeters), as well as the average error in azimuth (in degrees), in elevation (in degrees), and in distance (in centimeters). The 3D error represented the distance in centimeters between the 3D position indicated in each trial by the participant and the actual 3D location of the sound source (the speaker). The other errors referred to the absolute difference in each dimension separately. Note that when computing the absolute error for one dimension (e.g., azimuth) trials were collapsed irrespective of the other dimensions (i.e., elevation and depth). For instance, when computing the mean absolute error with respect to the position at +40 degrees we averaged the absolute error for speakers at ± 20 degrees in elevation and 35/55 cm in depth. Each dependent variable was computed separately for the two listening conditions.

To study head behavior, we focused on two dependent variables: the number of movements (identified as peaks of velocity) and the extent of head rotation around the vertical axis (head turning toward right-left) and the horizontal axis (head turning toward up-down). To calculate these variables, we considered all head rotations occurring within a trial. For instance, if the head rotated 20° to the right and 40° to the left, the amplitude for that trial was calculated as the sum of the two, therefore 60° and the same has been done for the rotation around the horizontal axis.

To study the head and hand movements, we used a custom-made software for the motion tracking analysis, running on MATLAB R2019b⁴.

All data were analyzed using Linear Mixed Effects (LME) or Generalized Linear Mixed Effects (GLME) models in R studio (Bunn & Korpela, 2019) with the packages lme4 (Bates, Mächler, Bolker, & Walker, 2015), car (Fox & Weisberg, 2021), and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017). The raw data can be retrieved from osf.io/57chk.

3.4. Results

3.4.1. Front-back discrimination task

To study the effect of listening conditions (active and static) on front-back discrimination performance, we entered the number of correct responses in a GLME model (family = binomial) with listening condition as categorical fixed effects and the intercepts of the participants as a random effect. The main effect of listening condition emerged ($X^2(1) = 58.21, p < 0.001$). As shown in Figure 13, moving the head during the presentation of the sound ($M = 0.92, DS = 0.27$) led to better performance compared to keeping the head fixed ($M = 0.61, DS = 0.49$). This effect emerged consistently for all participants but one.

In the active listening condition, participants made an average of 3.79 head-movements per trial ($SD = 1.96$) with an average amplitude of 23.94° ($SD = 11.73^\circ$). The extent of their head rotation around the vertical axis was on average 46.99° ($SD = 23.00^\circ$) and 20.05° around the horizontal one ($SD = 12.86^\circ$).

⁴ We calculated the tangential velocity on the x, y, and z axis (expressed in degrees of rotation) using a two-point central difference derivative algorithm (Bahill & McDonald, 1983) with five points for the half-window for the head movements and thirty points for the hand movements. The onset and the end of both the movements were computed automatically using a velocity-based threshold (10°/s) and were then checked manually by visualizing the spatial rotation changes of the head and the hand and their speed. This procedure served to establish the spatio-temporal profile of head and hand behavior and extract relevant parameters for subsequent analyses. It also served to reject all trials in which participants did not comply with the instructions (i.e., they made anticipatory hand movement during sound playback) or because of artefacts or lack of data. Head movements below 10° were discarded.

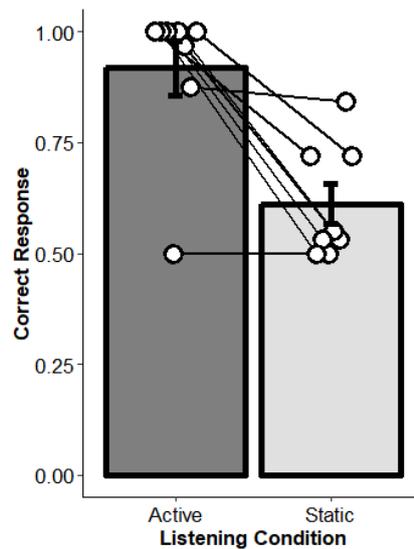


Figure 13 STUDY 2 - FRONT-BACK DISCRIMINATION PERFORMANCE

Effects of listening conditions on front-back discrimination performance number of correctly reported responses were reported. Individual participants are identified by points connected by a line. The overall mean is shown as bar plot, with error bars indicating standard errors.

3.4.2. Sound localization task in the frontal plane

Similar LME models were used to analyze the sound localization performance in the frontal plane, applied separately to 3D, azimuth, elevation, and distance errors. Moving the head during the sound presentation resulted in smaller 3D errors compared to the head fixed condition, as revealed by the significant main effect of listening condition ($X^2(1) = 18.42, p < 0.001$). As can be seen from the Figure 14, active listening led to better performance for all participants except one.

The same advantage for the active condition was observed when analyzing the azimuth error ($X^2(1) = 51.21, p < 0.001$), and the elevation error ($X^2(1) = 7.07, p < 0.008$), but it did not emerge for distance errors ($X^2(1) = 1.19, p = 0.28$). For more detailed information, see Table 4.

In the active listening condition, participants made an average of 2.64 movements per trial (SD = 1.24), with an average amplitude of 20.94° (SD = 19.36°). They rotated the head on average 36.47° to the left-right (SD = 22.83°) and 23.02° around the horizontal axis (SD = 13.62°).

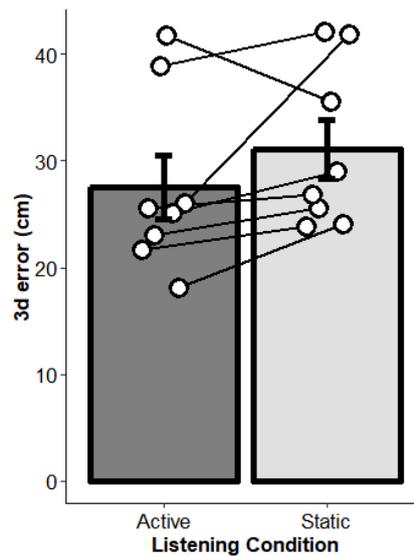


Figure 14 *STUDY 2 - SOUND LOCALIZATION PERFORMANCE*

Effects of listening conditions on 3d errors. differences in centimeters between the position chosen by the participant and the actual position of the source that considers the three dimensions of the space were reported. Individual participants are identified by points connected by a line. The overall mean is shown as bar plot, with error bars indicating standard errors.

	3D error (deg)		Azimuth error (deg)		Elevation error (deg)		Distance error (deg)	
	Active	Static	Active	Static	Active	Static	Active	Static
Mean	27.54	31.09	14.61	21.14	24.48	27.70	12.45	13.14
(SD)	(8.33)	(7.67)	(6.91)	(11.40)	(13.30)	(12.82)	(2.89)	(1.53)

Table 4 *STUDY 2 - ACTIVE AND STATIC ERROR*
Active and static 3D error, azimuth, elevation, and distance mean errors.

3.5. Discussion

In the present study we examined if letting participant move the head can improve sound localization as compared to a static condition in ageing adults with presbycusis. We focused on individuals with asymmetrical hearing loss, a situation in which there is an imbalance between the binaural cues (which reduce localization in the horizontal plane) and less possibility of exploiting the monaural cues (which reduce front-back disambiguation and the ability to discriminate between different elevations). Our findings clearly show that the advantage for spatial hearing of spontaneous head-movements during sounds as compared to a static condition was detected both for front-back discrimination of sound position and for localization of sounds in azimuth and elevation in front space.

Several previous works have investigated spatial hearing in ageing participants with presbycusis, documenting decreased or impaired performance along several spatial dimensions (Abel et al., 2000; Middlebrooks, 2015; Middlebrooks & Green, 1991; Neher, Laugesen, Sogaard Jensen, & Kragelund, 2011; Noble et al., 1994; Rakerd et al., 1998). The mean error degree in the static listening condition in our study (21.1° in azimuth) is in line with what has been found by Neher and colleagues (2011). However, participants in all these studies were always tested with the heads held still. In fact, head-supports were sometimes employed to provide “a physical reminder not to move the head during the trial” (Rakerd et al., 1998, p.468). To the best of our knowledge, there is only one exception to this wide-spread static-head approach. Otte and colleagues (2013) assessed the ability of old and young adults to localize sound sources that changed in both azimuth and elevation using head-orienting movements. Target sounds were set to last 150 ms, precisely to ensure that the head-saccades toward the sound “always started after stimulus offset, which denied listeners potential acoustic feedback during their response” (Otte, Agterberg, Van Wanrooij, Snik, & Van Opstal, 2013, p. 264). Yet, participants were asked to move their head during the task, thus they encoded sound position within a reference-frame that served head-movement. Although the effect of head-movements was not directly tested (comparing an active listening condition with a static one, as here) and despite the sound was not long enough to allow dynamic listening, it is interesting to note that using the head as pointer led to comparable localization performance in young and older adults. This result was specific to the horizontal plane, whereas on the vertical plane sound discrimination was again more affected in older adults compared to younger controls. To date, however, it remained to be ascertained if older adults could exploit head-movements during sound to improve sound localization, which we report here for both front-back discrimination and sound localization in the frontal plane for azimuth and elevation.

The effect of head movement in our study emerged clearly. Front-back discrimination improved from 61% accuracy in the static listening condition to 92 % accuracy when the head was free to move. Likewise, the 3D error decreased from 31.0 to 27.5 cm. Interestingly, this reduction occurred significantly for the horizontal and the vertical dimension, but not for depth (see Table 4). The improvement in the vertical dimension is particularly noteworthy, as this is exactly the axis for which spatial hearing difficulties have been repeatedly documented. In sum, at the net of identical auditory cues and central auditory processing skills, all participants except one were capable of exploiting head-movements when aiming to localize the sound. These findings are consistent with the literature on the benefit of head movements on sound localization in young subjects with normal hearing or hearing loss (Blauert, 1969; Brimijoin & Akeroyd, 2012; Iwaya et al., 2003; Macpherson, 2013; Makous & Middlebrooks, 1990; McAnally & Martin, 2014; Morikawa, Toyoda, & Hirahara, 2013; Perrett & Noble, 1997b; Pollack & Rose, 1967; Thurlow & Runge, 1967; Vliegen, Van Grootel, & Van Opstal, 2004; Wallach, 1940; Wightman & Kistler, 1999).

There are two reasons subtending the benefit of head movement while listening: on one hand this advantage is related to the possibility of extracting richer auditory cues. Dynamic cues can reduce ambiguity since moving the head introduces extra binaural and spectral cues enhancing localization (Kato et al., 2003; Perrett & Noble, 1997b). Head movements provide an accumulation of head-related transfer functions (HRTFs), refining localization and shrinking the cone of confusion. They also alter the amplitude and phase of the sound waves reaching ears, providing dynamic binaural cues (Risoud et al., 2018). Those can compensate for the disruption of monaural spectral cues (Kato et al., 2003; Perrett & Noble, 1997b), as in the case of presbycusis, which is accompanied by a hearing loss for high frequencies. On the other hand, letting the participants free to move their head and listen actively fosters their engagement and their implementation of head-related spontaneous strategies which represent a crucial cognitive resource when tackling a sound localization task, especially when auditory cues are altered (Coudert et al., 2021; Valzolgher, Alzhaler, et al., 2020; Valzolgher, Verdelet, et al., 2020).

In conclusion, this study suggests that allowing head movements during listening can improve auditory spatial skills in elderly people with presbycusis. Our findings highlight the need to exploit more ecological approaches when measuring spatial hearing abilities in this ageing population. Unlike previous studies (Iwaya et al., 2003; Perrett & Noble, 1997b), we studied participants in a common listening experience, leaving them to move freely and naturally and we provided enough time to implement head movements (5 seconds sounds). These results also open new directions for training spatial abilities in ageing people with presbycusis. Although in the present work head movements were most likely implemented spontaneously, active head-orienting to sounds can be trained. Recent evidence in this direction has emerged in normal-hearing young adults with one ear plugged (Valzolgher, Verdelet, et al., 2020). Future studies could test training paradigms to promote effective head-orienting strategies during sound localization also in ageing people with presbycusis.

Chapter 4. **Study 3: Reaching to sounds in virtual reality: A multisensory-motor approach to promote adaptation to altered auditory cues**

This study has been published in *Neuropsychologia* in 2020.

The following contents are reported from the same journal article.

Valzolgher, C., Verdelet, G., Salemme, R., Lombardi, L., Gaveau, V., Farné, A., & Pavani, F. (2020). Reaching to sounds in virtual reality: A multisensory-motor approach to promote adaptation to altered auditory cues. *Neuropsychologia*, 149, 107665. <https://doi.org/10.1016/j.neuropsychologia.2020.107665>

4.1. Abstract

When localizing sounds in space the brain relies on internal models that specify the correspondence between the auditory input reaching the ears, initial head-position and coordinates in external space. These models can be updated throughout life, setting the basis for re-learning spatial hearing abilities in adulthood. In addition, strategic behavioral adjustments allow people to quickly adapt to atypical listening situations. Until recently, the potential role of dynamic listening, involving head-movements or reaching to sounds, have remained largely overlooked. Here, we exploited visual virtual reality (VR) and real-time kinematic tracking, to study the role of active multisensory-motor interactions when hearing individuals adapt to altered binaural cues (one ear plugged and muffed). Participants were immersed in a VR scenario showing 17 virtual speakers at ear-level. In each trial, they heard a sound delivered from a real speaker aligned with one of the virtual ones and were instructed to either reach-to-touch the perceived sound source (Reaching group), or read the label associated with the speaker (Naming group). Participants were free to move their heads during the task and received audio-visual feedback on their performance. Most importantly, they performed the task under binaural or monaural listening. Results show that both groups adapted rapidly to monaural listening, improving sound localization performance across trials and changing their head-movement behavior. Reaching the sounds induced faster and larger sound localization improvements, compared to just naming its position. This benefit was linked to progressively wider head-movements to explore auditory space, selectively in the Reaching group. In conclusion, reaching to sounds in an immersive visual VR context proved most effective for adapting to altered binaural listening. Head-movements played an important role in adaptation, pointing to the importance of dynamic listening when implementing training protocols for improving spatial hearing.

4.2. Introduction

Spatial hearing is a remarkable ability of the brain. To determine the spatial coordinates of sounds in the environment, the cognitive system exploits auditory cues resulting from the interactions between sound waves, the head and outer ears (Middlebrooks, 2015; Middlebrooks & Green, 1991; Wallach, 1940). Auditory cues are distinguished in binaural cues, which involve a combined computation of the signals reaching the two ears and are used for sound localization in the horizontal dimension (Rayleigh, 1907), and monaural cues which involve the information available at each single ear and are crucial to localize sound in elevation, depth and anterior-posterior axis (Angell & Fite, 1901). In normal hearing listening condition, both monaural and binaural cues are weighted and combined to achieve optimal localization of auditory events in the 3D space (Carlile et al., 2005; see also Shinn-Cunningham et al., 2000). Auditory cues change during the life span, due anatomical modifications of one's own body during development (Clifton et al., 1988) or changes in hearing threshold with ageing (Cranford, Andres, Piatz, & Reissig,

1993; Dobрева, O'Neill, & Paige, 2011). In addition, they change on a daily basis, due to the diversity of listening contexts to which we are all exposed (Majdak et al., 2010). It is now clear that the brain remains capable to cope with these continuous changes in auditory cues throughout life (Carlile, 2014; Keating & King, 2015; Rabini et al., 2019; Strelnikov et al., 2011; Valzolgher, Campus, et al., 2020; Van Wanrooij & Van Opstal, 2005). In the present work, we examined reaching to sounds in virtual reality as a multisensory-motor strategy for adapting sound localization abilities in adulthood.

Updating of sound-space correspondences typically occurs in a multisensory environment. Whenever the other sensory systems (e.g., vision) provide reliable spatial information about auditory events, the brain exploits these additional sensory sources to calibrate and optimize internal models for spatial hearing (Keating & King, 2015). This notion emerged from studies that examined re-learning of sound-space correspondences when binaural auditory cues were temporarily altered using monaural ear-plugs (Rabini et al., 2019; Strelnikov et al., 2011; Trapeau & Schönwiesner, 2015; Valzolgher, Campus, et al., 2020), or monaural auditory cues were modified through ear molds (Van Wanrooij & Van Opstal, 2005) or non-individualized HRTFs (Head-Related Transfer Functions; (Hendrickx et al., 2017; Honda et al., 2007; Parseihian, Jouffrais, & Katz, 2014; Parseihian & Katz, 2012; Steadman, Kim, Lestang, Goodman, & Picinali, 2019a; Stitt, Picinali, & Katz, 2019). In these simulated altered-listening conditions, multisensory training procedures proved effective for adapting to novel auditory cues (for reviews see: Carlile, 2014; Irving & Moore, 2011; Keating & King, 2015; Knudsen & Knudsen, 1985; Mendonça, 2014). For instance, Strelnikov et al. (2011) studied the effects of audio-visual vs. auditory-only training on sound localization in monaurally-plugged adults. They found that performance improvements were larger after a training that exploited spatially and temporally congruent audio-visual inputs, compared to a training based on auditory information alone (Strelnikov et al., 2011). These findings complement other research in which participants received feedback after their sound localization response, through audio-visual (Shinn-Cunningham et al., 1998; Zahorik et al., 2006, see also Majdak et al., 2013) or visual cues (Bauer, Matuzsa, Blackmer, & Glucksberg, 1966; Kumpik, Kacelnik, & King, 2010; Mendonça et al., 2013). Finally, converging evidence in favor of multisensory-based training emerged from research in ferrets with unilateral and bilateral cochlear implants. Performance in early-deafened ferrets with bilateral cochlear implants improved consistently after a multisensory training that used interleaved auditory and visual stimuli (Isaiah et al., 2014).

In natural, off-laboratory conditions, spatial hearing is a dynamic and sensorimotor task. Listeners spontaneously orient their head, eyes and trunk to the auditory source (e.g., when turning to listen to a speaker behind us), or they move their body in the environment to approach the auditory source (e.g., when reaching to grasp the mobile phone ringing on the table). These dynamic and sensorimotor interactions with sound sources are almost ubiquitous in our everyday listening conditions. Head-orienting

to sounds is a spontaneous behavior: it is detected in infants 2-4 days after birth (Muir & Field, 1979) and it follows the shortest path to the sound already between 6 and 9 months of age (van der Meer et al., 2008). Although adults have different propensities for head-movement (Fuller, 1992), they can easily orient their head to sounds when required by the task (Brimijoin, McShefferty, & Akeroyd, 2010). Experimental evidence shows that head-movements during listening improve sound localization (Noble, 1981; Thurlow & Runge, 1967), and head-movement strategies change in altered hearing conditions in the attempt of coping with the hearing difficulty. For instance, Brimijoin et al. showed that people with hearing-impairment have more complex head-movements (Brimijoin et al., 2010). Also, people with asymmetrical hearing impairment use head-orienting to maximize the level of a target sentence in their better ear when solving a speech-in-noise task (Brimijoin & Akeroyd, 2012). Yet, studies that examined adaptation to altered auditory cues have typically limited the possibilities for head-movements using a chinrest (Kumpik et al., 2010; Strelnikov et al., 2011) or by instructing participants to refrain from head-movements during listening (Rabini et al., 2019; Van Wanrooij & Van Opstal, 2005; Zahorik et al., 2006).

However, notable exceptions to the static listening approach have emerged in the last decade. Using gamified scenarios, researches have asked participants to use a hand-held tool to hit a moving sound presented in virtual auditory space (Honda et al., 2007; Ohuchi, Iwaya, Suzuki, & Munekata, 2005), or to shoot audio-visual moving targets presented in virtual reality (Poirier-Quinot & Katz, 2018). In all these cases participants were free to move their heads and bodies. In more controlled conditions, Parseihian and Katz (2012) studied how people could adapt to non-individualized head-related transfer functions (HRTFs) when localizing virtual sound sources (Parseihian & Katz, 2012). They explored the effects of a multi-modal training platform in which participants were involved in an active game-like scenario. They actively searched for animal sounds, scanning the space around them (front and back), using a hand-held track-ball. Again, participants were free to move their heads and hands. This training improved vertical localization performance and reduced front/back confusions, showing that people could rapidly adapt to novel monaural cues. Using an identical training approach, but with longer sessions and with perceptually worst-rated non-individualized HRTFs, Stitt et al. (2019) also found evidence for some degree of adaptation (Stitt et al., 2019). Another recent study combined gamification, virtual reality and free head-movements in a training protocol aimed at ameliorating sound localization with virtual sounds (Steadman et al., 2019). These authors also trained participants to adapt to non-individualized HRTFs, hence to new monaural cues. They asked participants to listen to stationary virtual sounds, either with the head static or free to move. Compared to the static listening condition, when the head was free to move sound localization improved even over a very short timescale, corroborating the importance of head movements when adapting to novel auditory cues.

Despite these studies provide initial evidence in support of a key role of dynamic and sensorimotor interactions with sounds when adapting to altered auditory cues, they did not investigate the qualitative and quantitative nature of the spontaneous head-movements during listening (e.g., head-movement number or amplitude, head-movement strategies). As such, they only provide limited evidence of the role of head-movements in adapting to novel listening conditions. Note that in the head-constrained context, researchers can only examine the cognitive mechanisms of adaptation, such as cue-remapping (i.e., changes in the internal models that specify the correspondence between auditory cues and spatial locations) or cue-reweighting (i.e., changes in the contributions each auditory cue in these internal models). By contrast, the conditions in which the head is free to move are informative about the strategies (implicit or explicit) that participants use to cope with the altered listening conditions. In addition, although some of the previous literature point to a joint contribution of head-movements and reaching movements in improving sound localization with altered auditory cues (Honda et al., 2007; Ohuchi et al., 2005; Parseihian & Katz, 2012), several questions remain open. Specifically, are reaching movements relevant in promoting the useful strategies for adapting to altered spatial hearing conditions? Could they trigger different head-movements during sounds compared to alternative and less-interacting responses (e.g., naming labels that identify the speaker position)? Finally, the vast majority of studies that examined the effects of dynamic and sensorimotor interactions with sounds when adapting to novel auditory cues focused on monaural cues. Thus, it remains an open question to what extent head-movements can also allow fast adaptation to altered listening when binaural rather than monaural cues are changed.

In the present study we addressed all these questions by asking two groups of participants to perform a sound localization task in an interactive visual virtual reality scenario, with real sounds presented within reaching distance. Participants wore a Head-Mounted Display (HMD) and perceived themselves in the center of an empty virtual room with visible virtual speakers. In each trial, a 3 s sound was emitted from a real speaker aligned with one of the visible virtual speakers. All participants were asked to identify from which virtual speaker the sound originated and were informed that they were free to move their head during sound emission to solve the task. Crucially, half of participants were instructed to stop sound emission by reaching to the correct source using a hand-held controller (Reaching group). If the participant reached to the correct speaker the sound stopped, promoting a sense of agency over the auditory change. If instead the participant reached towards the wrong speaker the sound continued playing and the correct speaker started flashing, providing multisensory cues to actual sound location until the participants gave the correct response (see Figure 15 B). The other half of participants received identical stimulation and identical feedback but were instructed to stop sound emission by naming the number that identified the active speaker (Naming group).

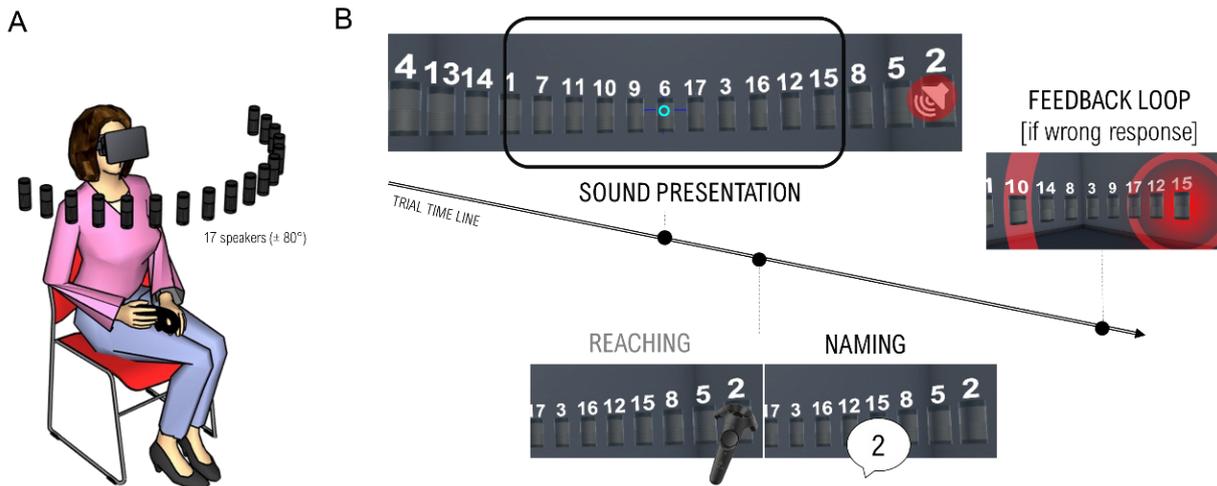


Figure 15 STUDY 3 - EXPERIMENTAL PROCEDURE AND SETTING

(A) Schematic representation of participant and speaker positions. (B) Schematic representation of 1 trial: participants saw the virtual room and part of the speakers' array (the black square represents participants starting view). They were allowed to turn their heads and move their eyes to see all speakers. In the Reaching group, participants stopped the sound by reaching and touching with the controller the virtual speaker emitting the sound. In the Naming group, the sound was stopped by the experimenter when participants read aloud the label above the speaker emitting the sound. If the answer was wrong, the correct source started to flash providing a visual feedback.

To examine the role of these two response instructions when adapting to altered spatial hearing conditions we focused specifically on binaural cues alteration, using plugging and muffing of the right ear. Unlike previous related works (Honda et al., 2007, 2013; Ohuchi et al., 2005; Parseihian & Katz, 2012; Stitt et al., 2019), in our study monaural cues at the unplugged left ear remained unaltered. Instead, we aimed to simulate conductive hearing loss. In terms of auditory cues, our manipulation clearly affected inter-aural level differences (ILD). In addition, because we adopted a slowly changing broad-band sound we made inter-aural time differences (ITD) largely ineffective (see Methods section).

If reaching to sounds play a role in adaptation to altered binaural cues, performance in the monaural listening condition should improve faster for the Reaching compared to the Naming group. The rationale for this hypothesis is linked to the fact that reaching to sounds requires a coordination of different effectors (eyes, head, hand) into a common reference frame (Cohen & Andersen, 2002). In turn, this may result in a more stable (or salient) spatial coding of sound source location and favor adaptation to altered spatial hearing conditions. In addition, reaching movement may help to direct attention towards the position occupied by the sound source, because of additional visual, proprioceptive and kinaesthetic cues resulting from the action.

4.3. Material and Methods

4.3.1. Participants

Twenty-eight participants (age: $M = 20.93$, $SD = 2.48$, range [18–30], 5 males, 14 right-handed in the Reaching Group, 13 right-handed in the Naming Group) were recruited to participate in the experiment, mostly among undergraduate students at the University of Trento. All participants signed an informed consent before starting the experiment, which was conducted according to the criteria of the Declaration of Helsinki (1964, amended in 2013) and approved by the ethical committee at the University of Trento (protocol: 2018-018). All had normal or corrected-to-normal vision and reported no movement deficit. Hearing threshold was measured using an audiometer (Grason Stadler GSI 17 Audiometer) for all participants, testing different frequencies (250, 500, 1000, 2000, 4000 Hz), on the right and left ear separately. All participants had an average threshold below 11.7 dB HL.

4.3.2. Apparatus and stimuli

Visual virtual reality (VR) and kinematic tracking was implemented using an HTC Vive System, comprising one head-mounted display (HMD, resolution: 1080×1200 px, Field Of View (FOV): 110° , Refresh rate: 90 Hz), 1 controller (which was used by participants to interact with sounds), 1 tracker (which was placed above the speaker to track its position in real time) and 2 lighthouse base stations (which served for regular scanning of the position of the controller and the tracker). Tracking precision and accuracy of the HTC Vive System is adequate for behavioral research purposes, as recently measured by our research group (Gaveau et al., 2020; Verdelet et al., 2020). Specifically, the HTC Vive has submillimeter precision (0.237 mm) and near centimeter accuracy (9.0 mm when trackers are static; 9.4 mm when trackers are dynamic). The HMD was equipped with an SMI eye-tracking system (250 Hz). All stimuli were controlled and delivered using a LDLC ZALMAN PC (OS: Windows 10 (10.0.0) 64bit; Graphic card: NVIDIA GeForce GTX 1060 6 GB; Processor: Intel Core i7-7700K, Quad-Core 4.2 GHz/4.5 GHz Turbo - Cache 8 Mo - TDP 95W) using Steam VR software and the development platform Unity.

The experiment was entirely run in a 4×3 meters ordinary room almost devoid of furniture, not treated for being anechoic and quiet. The subject sat in the center of the room, with no constraints for her head (no chin-rest was used). The virtual environment in which participants were immersed was an ordinary squared room (grey wall), similar to the real experimental room. The room was empty and there were not objects on the walls, except a door which was placed in the back wall. In front of participants 17 visible speakers were positioned at ear level. In each session, subjects wearing the HDM saw themselves in the virtual scenario, sitting 50 cm from the center of a semicircle of 17 speakers, spanning about $\pm 80^\circ$ of visual angle. Each speaker was marked with a numeric label (from 1 to 17) that changed in each trial.

Auditory stimuli were delivered by a real, unseen speaker (JBL GO Black), which was moved by the experimenter and whose position was tracked in space. The auditory stimulus was an amplitude-modulated (4 Hz) white noise bursts (about 60 dB SPL as measured by a Decibel Meter TES1350A placed at ears level).

This solution allowed us the possibility of tracking the position of the tracker (i.e., the real speaker), the controller (handled by the participant) and the Head Mounted Display, via sensors on the headset and VR controllers using IR led positioned on the base stations (frequency sample 250 Hz). We design the software to allow the experimenter to track and align the real loudspeaker (the sound source) with the desired position in the virtual environment (i.e., in correspondence of one of the visible loudspeakers). Specifically, in each experimental trial, the experimenter moved the loudspeaker to the desired position in 3D space, following visual instructions concerning the pre-determined speaker position. These instructions were visible only to the experimenter, and they were delivered using the stimulus visualization monitor placed in the testing room. The monitor provided a bird-eye view of the experimental room and conveyed the pre-determined position of the sound source for that trial and the actual real-time position of the speaker. Using these visual instructions, the experimenter reached for computer-determined position rapidly, keeping the speaker approximately at ear-level. The spanning of about $\pm 80^\circ$ permitted to the experimenter to stand still in front of the participant and move the speaker silently stretching her right or left arms. Although the noise produced by the experimenter were minimal, participants were explicitly informed to pay attention to the target sounds, as any other sound in the room could in fact be deceiving about target sound position. Before starting, the experimenter showed how she can stay to the right of participants while providing sound from the left of participants by stretching her arm of the hand with the speaker. This was a demonstration that background noise could provide erroneous cues and thus discourage their use by participants, which were invited to focus to the specific target sound. Most importantly, pilot work in our laboratory showed that participants cannot reliably point to the speaker if the same placement procedure is used but no target sound is delivered (Gaveau et al., 2020).

To simulate monaural hearing experience, the left ear of each participant was occluded using an ear-plug (3M PP 01 002; attenuation value for high frequencies = 30 dB SPL; attenuation value for medium frequencies = 24 dB SPL; attenuation value for low frequencies = 22 dB SPL; Single Number Rating = 32 dB SPL; as reported by the manufacturer). In addition, a unilateral ear muff covered the right ear (3M 1445 modified to cover only the right ear; attenuation value for high frequencies = 32 dB SPL; attenuation value for medium frequencies = 29 dB SPL; attenuation value for low frequencies = 23 dB SPL; Single Number Rating = 28 dB SPL; as reported by the manufacturer). The combined use of an ear-plug and muff at the right ear clearly affected ILDs. In addition, by attenuating the intensity of our amplitude-modulated broadband signal to one ear we most likely reduced the available ITD cues. This

because we reduced the possibility of comparing the fine structure at low frequency, as well as the amplitude modulation signals that reach the two ears. Thus, we altered both binaural cues while leaving intact the possibility of exploiting monaural cues at the unplugged left ear.

4.3.3. Procedure

We measured auditory performance during five different blocks, under two different listening conditions: binaural listening (B) and simulated monaural listening (M) in which the left ear was plugged. In each block, participants performed 51 trials (3 trials for each source) of a sound localization task. First, participants were invited to sit down on a chair and wear the Head Mounted Display (HMD) and to perform an eye-tracker calibration (they had to follow a white dot moving on the screen with eyes). Then, they were immersed in the virtual scenario. Around them, the cylindrical speakers were located at the height of their ears. Participants were instructed that the starting position consisted in gazing at the central speaker (marked with a cross) and refrain from moving their head until the beginning of the sound. The system delivered the sound only when head (HMD) and eyes (tracked) were directed toward the center. In such a way, we obtained replicable head and eyes posture across trials and participants. As soon as the sound started, participants were free to move their eyes and head as they wanted.

After each block, participants were given breaks and were asked to judge their performance (“How do you judge your performance?” from 1 = really bad to 9 = really good) and their perceived effort during the block (“How much effort (in terms of energy and cognitive resources) did the task require?” from 1 = none to 9 = many). At the end of 5 blocks, participants were invited to complete a questionnaire presented using Google Forms. The questionnaire comprised general questions about personal data (i.e., age, gender), questions to investigate their relationship with sound during life, specific questions to deepen the quality of our VR technology and its ergonomics, a question concerning their strategy to perform the task, two questions concerning the feeling of being able to switch the sound off (Agency) and 15 items of 2 scales (Focus of Attention and Satisfaction), used typically to measure the engagement in video game-based environments. We adapted these items from the work of Wiebe et al. (2014) to make them more suitable for our VR experience (Wiebe, Lamb, Hardy, & Sharek, 2014). Participants indicated their agreement with each item using a 5-points Likert scale (from 1 = totally disagree to 5 = totally agree).

We divided participants in 2 groups: the Reaching group and the Naming group (Figure 15 B). The Reaching group performed the task by moving a controller with their dominant hand to touch the speaker emitting the sound. The Naming group performed the same task by reading aloud the number located above the presumed source speaker (see Figure 15 B). If the response was correct, the sound immediately stopped. If the response was wrong, the correct speaker produced a pulsing red light while the sound continued. This provided participants with audio-visual feedback about the correct position of

the sound source. Participants either moved the controller in contact with the correct source or named the number of the correct speaker. When the correct speaker was reached (or named) the sound stopped. Importantly, participants were informed that they were free to exploit head-movements while listening to target sounds, without time pressure on their response. Both reaching and naming responses were followed by a brief vibration of the hand-held controller (which was kept in hand also during the naming task). When participants in the Naming group were tested, a second experimenter in the room promptly registered the responses on the computer. Specifically, the second experimenter pressed a key on an additional VIVE controller to specify whether the response was correct or wrong, and at the end of the trial entered the spoken labels in the computer.

4.3.4. Head-movement pre-processing

To study head-movements, we calculated the tangential velocity on the x, y, z axis (expressed in degrees of rotation) using two-points central difference derivate algorithm (Bahill & McDonald, 1983) with 5 points for the half-window. The onset and the end of the movements were computed automatically using a velocity-based threshold ($10^\circ/s$). Head-movements were checked manually by visualizing the spatial rotation changes of the head and its speed using a home-made tool box in MATLAB R2018b.

4.4. Results

4.4.1. Sound localization performance

4.4.1.1. Immediate ear-plug effect

To assess the immediate effects of monaural plugging, we compared absolute and signed errors before and after ear-plugging. Absolute and signed errors (calculated in each trial as the absolute or signed difference in degrees between actual and reported azimuth position of the sound, respectively) were entered into separate Analysis of Variance (ANOVA) with TARGET POSITION (17 positions, from -80° to $+80^\circ$) and LISTENING CONDITION (binaural B, monaural M1) as within-participants variables and GROUP (reaching, naming) as between-participants variable. All statistical analyses using the software JASP 0.9.1.0 and R-studio (version 1.0.143).

The analysis on absolute errors revealed a main effect of LISTENING CONDITION, $F(1,26) = 64.79, p < .001, \eta^2 = 0.71$, a main effect of TARGET POSITION, $F(3.95, 102.60) = 5.96, p < .001, \eta^2 = 0.18$, and the 2-way interaction between LISTENING CONDITION and TARGET POSITION, $F(3.9, 101.51) = 7.36, p < .001, \eta^2 = 0.22$. Sound localization error increased significantly in M1 compared to the B listening condition (see black and light grey lines in Figure 16 A and B), particularly for target delivered more towards the plugged ear (left). This effect emerged irrespective of GROUP (main effect and all interactions,

p-values > .27). Likewise, the analysis on signed errors revealed a main effect of TARGET POSITION, $F(4.69, 121.90) = 9.42$, $p < .001$, $\eta^2 = 0.26$, and the 2-way interaction between LISTENING CONDITION and TARGET POSITION, $F(4.33, 112.64) = 8.08$, $p < .001$, $\eta^2 = 0.23$. Signed errors increased particularly for target delivered more towards the plugged ear (left). For signed errors we also found a main effect GROUP, $F(1,26) = 6.45$, $p = .02$, $\eta^2 = 0.20$ and the interaction between GROUP and LISTENING CONDITION ($F(1,26) = 4.36$, $p = .05$, $\eta^2 = 0.26$). Simple main effects revealed that the signed error of the Naming group was significantly impacted by listening condition ($F = 4.9$, $p = .05$), whereas this did not occur in the Reaching group ($F = 0.26$, $p = .62$). As shown in Figure 16 B, during monaural listening the Naming group localized sound more to the unplugged side (Binaural: $M = -0.39$; $SD = 3.14$; Monaural: $M = 4.36$; $SD = 16.53$), whereas the Reaching group did not present this bias (Binaural: $M = -0.88$; $SD = 3.55$; Monaural: $M = -1.69$; $SD = 14.34$).

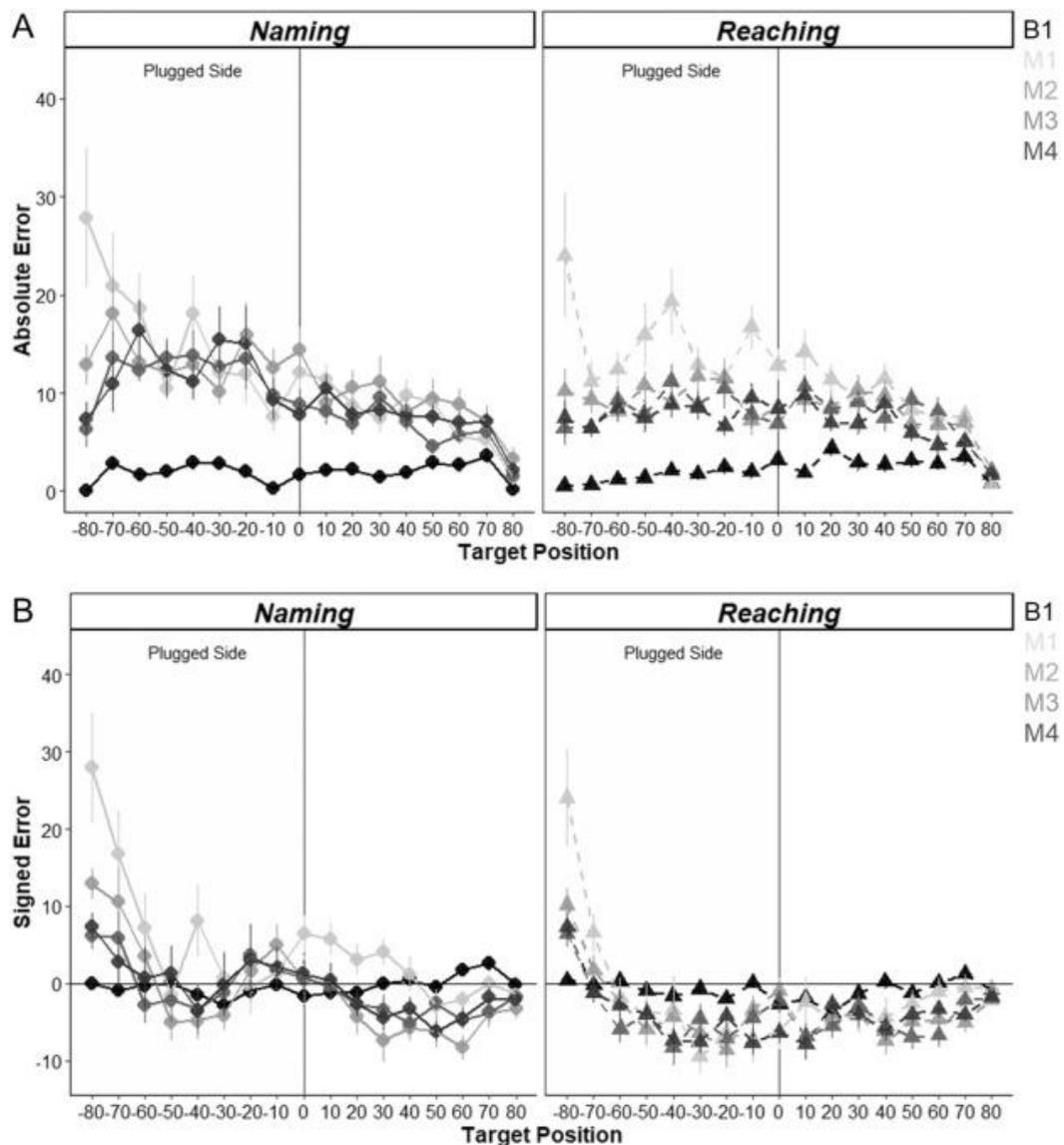


Figure 16 STUDY 3 - SOUND LOCALIZATION PERFORMANCE

Absolute (A) and signed (B) errors (in degrees) as a function of Blocks (B = binaural block, M1 = first monaural block, M2 = second monaural block, M3 = third monaural block, M4 = fourth monaural block) and group (Naming: left panel, circles; Reaching: right panel, triangles). Error bars represent the standard error of the mean.

Participants were aware of their performance change after monaural plugging. Performance assessment scores (“*How do you judge your performance?*” from 1 = really bad to 9 = really good) were lower in M1 (Reaching: $M = 3.29$; $DS = 1.90$; Naming: $M = 3.50$, $DS = 1.40$) compared to the B listening condition (Reaching: $M = 6.86$; $DS = 1.35$; Naming: $M = 6.57$, $DS = 0.85$; Chi-square = 24.14, $p < .001$, in non-parametric Friedman test), irrespective of GROUP. Participants also reported that monaural plugging required additional cognitive effort. Effort assessment scores (“*How much effort (in terms of energy and cognitive resourced used) did the task require?*” from 1 = none to 9 = many) were larger in M1 (Reaching: $M = 6.71$; $DS = 1.33$; Naming: $M = 6.21$, $DS = 1.72$) compared to the B listening condition (Reaching: $M = 3.29$; $DS = 1.49$; Naming: $M = 3.57$, $DS = 1.95$; Chi-square = 28, $p < .001$, in non-parametric Friedman test), again irrespective of GROUP.

4.4.1.2. Adaptation to monaural listening

To study adaptation to monaural listening across successive monaural blocks, we entered absolute and signed errors in separate ANOVAs with TARGET POSITION (17 positions, from -80° to $+80^\circ$) and MONAURAL BLOCKS NUMBER (M1, M2, M3 and M4) as within-participants variables, and GROUP (reaching, naming) as between-participants variable. The analysis on absolute errors revealed a main effect of MONAURAL BLOCK NUMBER, $F(3,78) = 8.51$, $p < .001$, $\eta^2 = 0.24$, a main effect of TARGET POSITION, $F(16, 416) = 8.89$, $p < .001$, $\eta^2 = 0.24$, and the 2-way interaction between MONAURAL BLOCK NUMBER and TARGET POSITION, $F(48, 1248) = 3.02$, $p < .001$, $\eta^2 = 0.10$. Participants improved across M blocks, especially for target positions more towards the plugged ear (left) (Figure 16 A and B). No main effect or interaction involving the GROUP variable emerged (all p -values $> .18$), when considering overall performance in a block. A similar analysis on signed errors revealed a main effect of MONAURAL BLOCK NUMBER, $F(3,78) = 5.252$, $p = .002$, $\eta^2 = 0.16$, a main effect of TARGET POSITION, $F(16, 416) = 10.68$, $p < .001$, $\eta^2 = 0.28$, and the 2-way interaction between monaural block number and target position, $F(48, 1248) = 2.64$, $p < .001$, $\eta^2 = 0.09$. Participants bias changed across M blocks, especially for target positions on the plugged side (left) (Figure 16 A and B). While participants perceived targets more toward the unplugged side at the beginning, repeating the task their started perceived the targets more toward the plugged side. Despite this pattern seems to differ in the 2 groups (the Naming group minimize the bias across blocks, while the Reaching group increased it toward the plugged side), no main effect or interaction involving the group variable emerged (all p -values $> .08$), when considering overall performance in a block.

To further investigate changes in localization error as a function of practice in the two groups, we analyzed performance changes across the 204 trials (51 trials in each of the 4 blocks) using a linear mixed model (LMM). Specifically, we entered the absolute error in a model with GROUP (naming, reaching) and PRACTICE (as a continuous variable of 204 items). Participant and target position were considered as random effects in the model. This analysis revealed that absolute errors were influenced

by PRACTICE ($t = -4.45, p < .001$), revealing that participants learned across monaural trials. Crucially, our model showed that error reduction was across trials was influenced by the type of task performed by participants. The Reaching group reduced localization errors more rapidly and to a greater extent compared to the Naming group ($t = -3.09, p = .002$) as shown in Figure 17 A. A similar analysis on signed error revealed that signed errors was generally influenced by GROUP ($t = -2.58, p = .01$). As shown in Figure 17 B, the Reaching group in general perceived the sound as shifted toward the plugged side, while the Naming group perceived it more toward the unplugged. Interestingly, both groups modified their bias by repeating the task ($t = -6.73, p < .001$), but differently ($t = 2.71, p = .007$). While the Naming group changed the direction of the bias across trial repetition, the Reaching group increased the bias toward the plugged side.

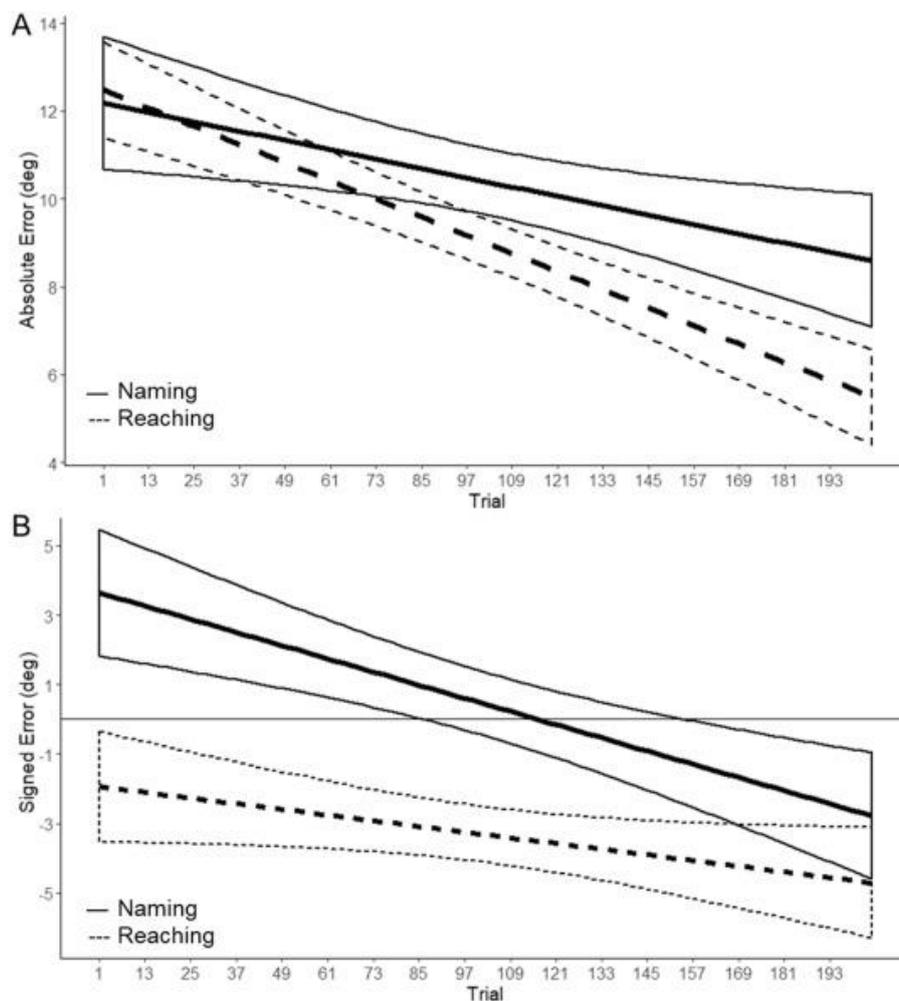


Figure 17 STUDY 3 - SOUND LOCALIZATION PERFORMANCE AS FUNCTION OF TRIAL
 Absolute (A) and signed (B) error as a function of trial for Naming (continuous line) and Reaching (dashed line) groups.

Participants were aware that their performance improved across monaural blocks. Performance self-assessment scores increased across monaural blocks (M4: Reaching: $M = 4.29$; $DS = 1.38$; Naming: $M = 3.93$, $DS = 1.69$; Chi-square = 10.02, $p < .001$, in non-parametric Friedman test), irrespective of GROUP. Participants did not report any change in terms of energy and cognitive resources used to do

94

the task across monaural blocks: effort assessment scores did not change across monaural blocks in either GROUP (all $ps > .23$).

4.4.2. Head movements

To study head-movement behavior, we measured the amplitude of the head rotation around the vertical axis for each detected movement. Next, we transformed the movement amplitude into a categorical variable with 10 bins (amplitude in each bin = 15.8° ; bins: 1 = 1.0° – 16.8° ; 2 = 16.8° – 32.7° ; 3 = 32.7° – 48.5° ; 4 = 48.5° – 64.3° ; 5 = 64.3° – 80.1° ; 6 = 80.2° – 96.0° ; 7 = 96.3° – 111.9° ; 8 = 113.9° – 126.8° ; 9 = 126.8° – 143.2° ; 10 = 143.2° – 159.3°) and counted the number of head movements in each amplitude bin, separately for each group and experimental block.

Furthermore, we considered two additional variables computed across all trials: head-rotation around vertical axis as a function of target position and overall head-direction bias during an entire block. The latter two indices were extracted by linear fitting of the changes in head-rotation extent as a function target position (see example in Figure 18). Head-rotation extent was calculated as the average between the leftmost and rightmost head-rotation, for any given target position. The slope of the fitting line captures head-rotation around the vertical axis: the larger the slope the more participants rotated their head as a function of target eccentricity. In the representative participant (ID = 21) shown in Figure 18, the slope remained stable from B to M1 (0.24 vs. 0.22) but increased from M1 to M4 (0.22 vs. 0.40), indicating a larger propensions to rotate the head towards peripheral targets in M4 compared to M1. The intercept of the fitting line captures the overall head-direction bias during an entire block, with negative values indicating a bias towards the left (plugged) side. In the same representative participant, the intercept became progressively more negative from B to M1 to M4 (-0.22 , -1.73 , -17.14 , respectively) revealing the bias of this participant to rotate the head further to the plugged side, thus bringing the unplugged ear towards the target.

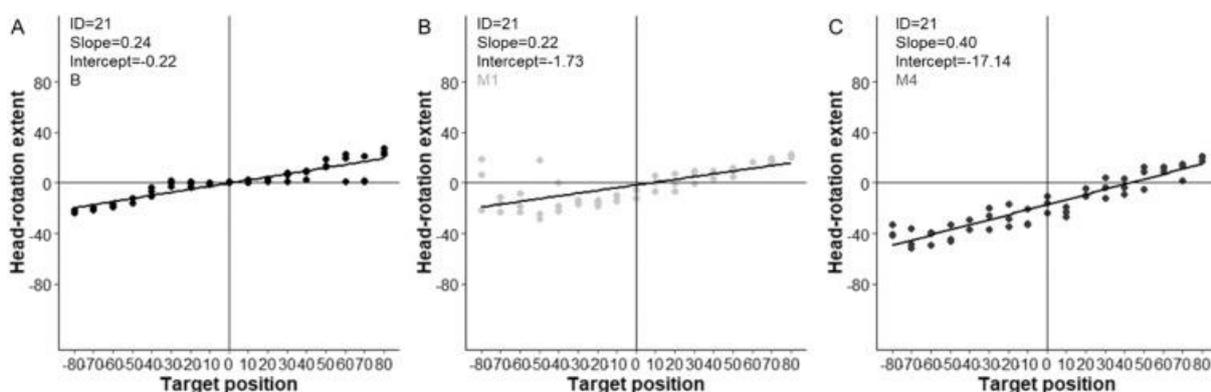


Figure 18 STUDY 3 - SLOPE AND INTERCEPT INDECES CALCULATION
Head-rotation extent as a function of target position of participants 21 during first binaural block B (A), the first monaural block M1 (B) and the last monaural block M4 (C).

4.4.2.1. Immediate ear-plug effect

We started by studying the immediate effect of monaural plugging on the number of head movements in the two groups, as a function of head-movement amplitude. To this aim, a Poisson regression model was fitted on the observed number of movements using `AMPLITUDE BIN` (covariate), `GROUP` (Naming, Reaching), `LISTENING CONDITION` (Binaural, Monaural), and all interaction terms. The participant was included as random effect. The deviance table⁵ for this model revealed a significant effect of `AMPLITUDE BIN` ($\chi^2(1) = 157.1710, p < .001$). Number of movements was inversely related with amplitude, as evident from the corresponding negative Poisson parameter estimate ($\beta_{\text{Amplitude}} = -0.58333, z = -12.537, p < .001$). Number of movements also increased significantly from the binaural (B1) to the first monaural (M1) listening block ($\beta_{m1} = 0.67464, z = 6.647, p < .001$) for both groups (main effect of Block, $\chi^2(1) = 44.1786, p < .001$). This increase concerned primarily head-movements of smaller amplitude (Amplitude by Block, $\chi^2(1) = 15.861, p < .001$). All the other effects were not statistically significant ($ps > 0.05$) (Figure 19 A).

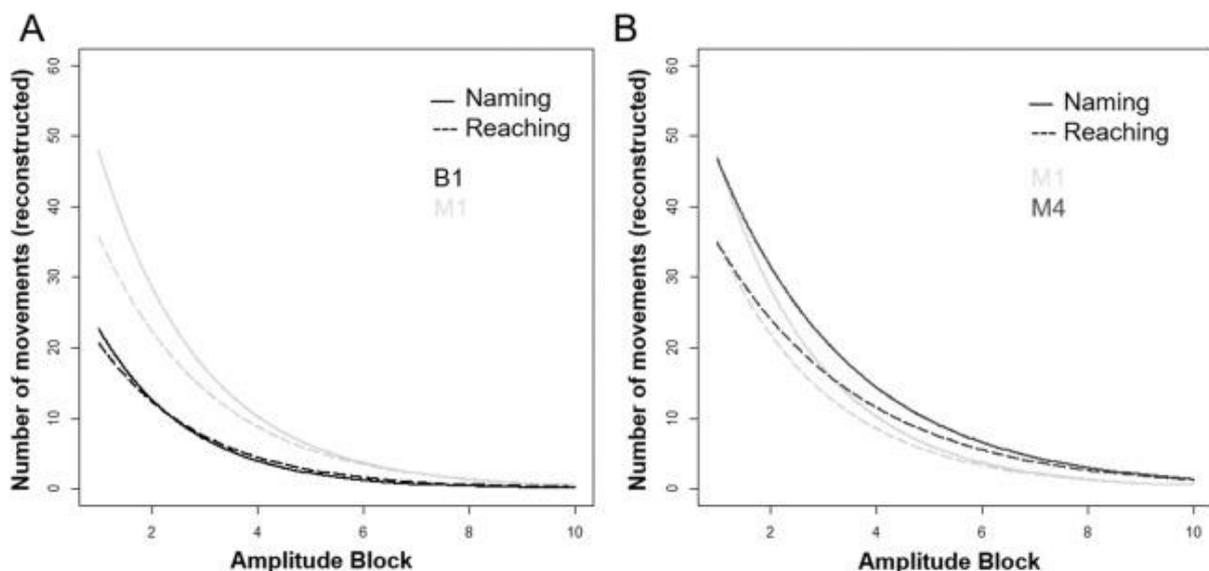


Figure 19 STUDY 3 - NUMBER OF MOVEMENT AND AMPLITUDE

A) Poisson model representation. N° of movements as a function of Amplitude block (data binning: 10 bins: 1 = 1.0°–16.8°; 2 = 16.8°–32.7°; 3 = 32.7°–48.5°; 4 = 48.5°–64.3°; 5 = 64.3°–80.1°; 6 = 80.2°–96.0°; 7 = 96.3°–111.9°; 8 = 113.9°–126.8°; 9 = 126.8°–143.2°; 10 = 143.2°–159.3°); Black patters correspond to binaural listening condition (B1) and light grey pattern correspond to monaural listening condition (M1). The continuous patterns represent the Naming group while dashed one the Reaching group. B) Poisson model representation. N° of movements as a function of Amplitude block. Light grey patters correspond to the first monaural block (M1) and dark grey pattern correspond to the fourth monaural listening condition (M4). The continuous patterns represent the Naming group while dashed one the Reaching group.

Next, we examined how participants explored auditory space with their head during sound emission, in the binaural and in the first monaural block. To this aim we focused on (1) head-rotation around vertical

⁵ Deviance decomposition table for the Poisson model derived using the `anova` function in `Car` package by Fox. We adopted the `glmer` function in the `lme4` package for parameter estimates and the `lmerTest` package for the computation of the corresponding p -values.

axis as a function of target position (slope of the fitting line, as specified above) and (2) overall head-direction bias during an entire block (intercept of the fitting line). Both measures were entered separately into an ANOVAs with listening condition (B, M1) as within-participants variable and group (reaching, naming) as between-participants variable. The analyses on head-rotation around vertical axis revealed that participants in the Reaching group turned their head as a function of target position to a greater extent compared to those in the Naming group (main effect of group, $F(1,26) = 5.29, p = .03, \eta^2 = 0.17$). In other words, they explored auditory space with their head to a greater extent (compare B and M1 in Figure 20 A). No main effect or interaction involving LISTENING CONDITION emerged, indicating that this group difference was already present in the binaural listening condition.

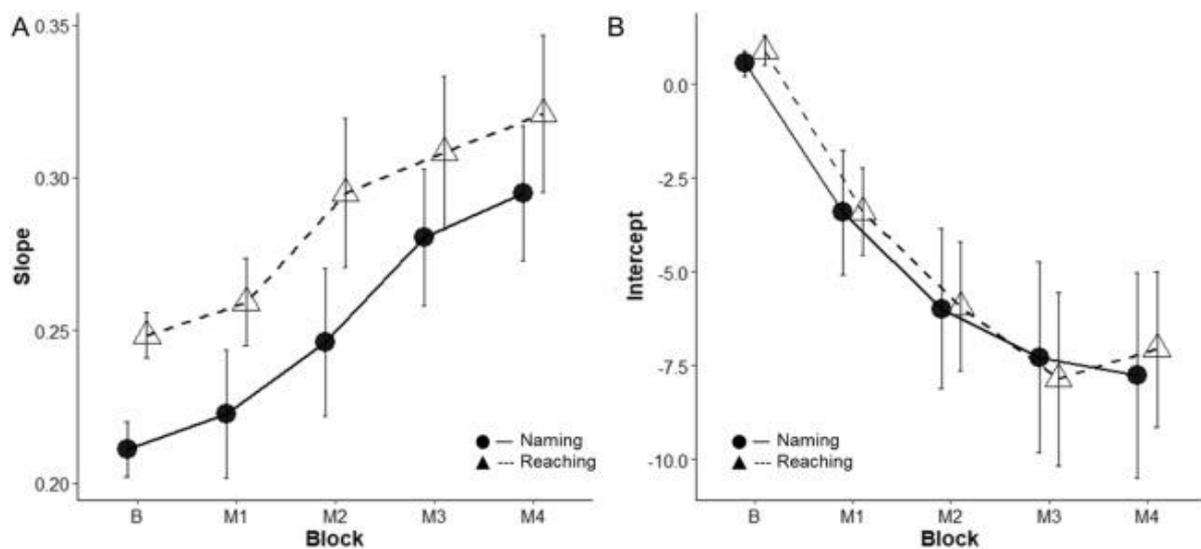


Figure 20 STUDY 3 - SPACE EXPLORED
Slope (A) and intercept (B) values as a function of block B1 M1 M2 M3 M4. Naming group (black circles) and Reaching group (white triangles).

The analysis on the overall head-direction bias revealed instead a main effect of LISTENING CONDITION, $F(1,26) = 19.09, p < .001, \eta^2 = 0.42$. Intercepts became negative in M1 (Naming: $M = -3.41, SD = 6.20$; Reaching: $M = -3.39, SD = 4.31$) compared to the B listening condition (Naming: $M = 0.56, SD = 1.25$; Reaching: $M = 0.90, SD = 1.50$). This means that, when listening with the left ear plugged, all participants turned their head leftward, to approach the sound with their unplugged (right) ear (compare B and M1 in Figure 20 B).

4.4.2.2. Adaptation to monaural listening

To study the effect of the adaptation to monaural listening on head-movements, we started by examining the number of head movements in the two groups, as a function of head-movement amplitude, in the first and in the last monaural block. As before, a Poisson regression model was fitted on observed number of movements as a function of AMPLITUDE BIN (covariate), GROUP (Naming, Reaching), BLOCK (M1, M4) with

all interaction terms. Participant number was included as random effect. The deviance table revealed a significant effect of AMPLITUDE BIN ($\chi^2(1) = 584.4407, p < .001$). Number of movements was inversely related with amplitude as evident from the corresponding negative Poisson parameter estimate ($\beta_{\text{Amplitude}} = -0.50891, z = -24.175, p < .001$). Number of movements also increased significantly from the first (M1) to the fourth (M4) monaural listening block, for larger amplitude levels ($\beta_{\text{Amplitude,M4}} = 0.11675, z = 4.321, p < .001$). This was also reflected from the significant interaction effect BLOCK by AMPLITUDE BIN ($\chi^2(1) = 18.6691, p < .001$). We also observed a significant effect for GROUP ($\chi^2(1) = 5.5511, p < .05$). The Reaching group was associated with a lower number of movements ($\beta_{\text{Reaching}} = -0.33190, z = -2.356, p < .05$). All the other effects were not statistically significant ($ps > 0.05$) (Figure 19 B).

Finally, we examined how participants explored auditory space with their head across monaural blocks. Slopes and intercepts across participants were entered into separate ANOVAs with MONAURAL BLOCK NUMBER (M1, M2, M3 and M4), as within-participants variable and GROUP (reaching, naming) as between-participants variable. Both analyses revealed only a main effect of MONAURAL BLOCK NUMBER (slope: $F(3,78) = 12.67, p < .001, \eta^2 = 0.32$; intercept: $F(1.57,40.86) = 9.47, p < .001, \eta^2 = 0.27$). As monaural blocks progressed, participants in both groups increased their exploration of auditory space and turned their head more to approach sounds with their unplugged (right) ear (Figure 20 A and B).

4.4.3. Relation between changes in sound localization performance and head-movements

4.4.3.1. Immediate ear-plug effect and head-movements

We first examined if the immediate ear-plug effect (i.e., changes in absolute localization error between M1 and B; with positive values indicating larger plug-effect) correlated with changes in the number of head movements (i.e., changes between M1 and B). When studied separately for each group, using Pearson correlation with Bonferroni correction, a significant relation emerged only for the Naming group (naming: $r = 0.57, p = .04$; reaching: $r = 0.12, p = .68$). However, this correlation was driven by a single participant ($r = 0.22, p = .46$ with the participant 18 was removed).

Next, we examined if the immediate ear-plug effect (i.e., changes in absolute localization error between M1 and B; positive values indicate larger plug-effect) correlated with changes in the space explored with the head (i.e., changes in slope between M1 and B). Unlike the case for number of head-movements described above, this correlation was reliable in each group (Pearson correlation with Bonferroni correction: reaching, $r = -0.66, p = .01$; naming, $r = -0.79, p < .001$; see Figure 21 A). Irrespective of group, the more participants increased explored space between B and M1 blocks, the less the cost of ear plugging on their sound localization performance.

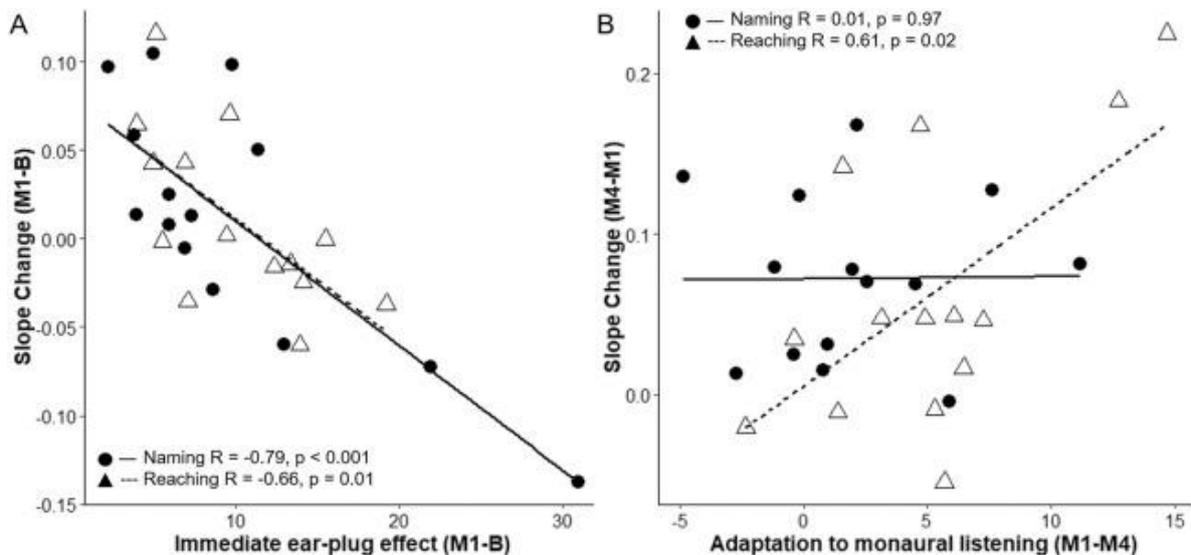


Figure 21 STUDY 3 - CORRELATION BETWEEN SPACE EXPLORED ROTATING THE HEAD AND PERFORMANCE
 A) Correlation between Ear-plug effect and Slope Change, B) Correlation between Slope Change and Adaptation to monaural listening. In both graphs participants in the Reaching group are indicated with white triangles (regression line shown as dashed) and those in the Naming group are indicated with black circles (regression line shown as continuous).

4.4.3.2. Adaptation to monaural listening and head-movements

The number of head movements also increased between M1 and M4 block (see section 4.4.2.2). We examined if adaptation (i.e., changes in absolute localization error between M4 and M1; positive values indicate larger improvement after trials repetition) correlated with the changes in the number of head movements (i.e., changes between M4 and M1). We did not find any relations for both groups (all p s > 0.31).

The portion of space explored with the head also changed across monaural blocks (comparison between M1-M4 blocks, see section 4.4.2.2). We examined if adaptation (i.e., changes in absolute localization error between M4 and M1; positive values indicate larger improvement after trials repetition) correlated with changes in the space explored with the head (i.e., changes in slope between M4 and M1). This correlation is shown in Figure 20 B, separately for each group (Pearson correlation with Bonferroni correction: reaching, $r = 0.60, p = .02$; naming, $r = 0.01, p = .97$). A positive relation emerged for the Reaching group, but not for the Naming group. Comparing correlations across group (as in Eid et al., 2011) produced a marginally significant between-group difference ($z = -1.602, p = .055$). This suggests that compensatory head-movement behavior may relate with adaptation more for participants in the Reaching group.

4.4.4. VR experience

At the end of the experiment, participant replied to questions about their experience using VR system, to assess also the feasibility of our approach. They judged that the overall experience with virtual reality was positive ($M = 7.86, SD = 1.15$, using a scale from 1 = negative to 9 = positive), that the scene appeared realistic ($M = 6.61, SD = 1.47$, using a scale from 1 = unrealistic to 9 = very realistic) and reported no

feeling of losing balance ($M = 1.53$, $SD = 1.07$, using a scale from 1 = not at all to 9 = completely) or being annoyed ($M = 2.61$, $SD = 1.81$, using a scale from 1 = not at all to 9 = completely).

Participants were also queried about the sense of agency and the perceived gaming experience when performing the two tasks (i.e., reaching or naming).

4.4.5. Sense of agency

Participants indicated their agreement with 2 items (“I felt that I can turn off the sound”, “I felt that I can act on the sound”) using a 5-points Likert scale (from 1 = totally disagree to 5 = totally agree). Participants in both groups felt equally that they could act on sounds (reaching: $M = 2.29$, $SD = 0.99$; naming: $M = 1.71$, $SD = 0.99$; $W = 64$, $p = .14$ on Mann Whitney). However, the Reaching group reported the experience of turn off the sound strongly ($M = 4.43$, $SD = 0.51$) compared to the Naming group ($M = 1.21$, $SD = 0.58$), $W = 0.00$, $p < .001$ (Mann Whitney).

4.4.6. Gaming experience

To investigate their gaming task-related experience, we proposed 15 items of 2 scale of the User Engagement Scale (UES) questionnaire, typically used to measure engagement during video-game playing (Wiebe et al., 2014): focus of attention scale and satisfaction scale (Figure 22). We compared responses of the two groups for each item of the 2 scales, as well as the cumulative indices (average of the respective items, Cronbach's α : 0.8; 0.89) in t-test analyses. Overall, our virtual reality scenario was an effective method to create a game-like situation, capable of actively involving participants. The satisfaction scale items revealed a difference between tasks: participants of the Reaching group ($M = 3.71$, $SD = 0.83$) perceived the experience more rewarding compared to the Naming group ($M = 2.93$, $SD = 0.83$; $W = 47.50$, $p = .015$). Moreover, they expressed more interest in doing the task (Reaching: $M = 4.71$, $SD = 0.47$; Naming: $M = 4.07$, $SD = 0.92$, $W = 60$, $p = .05$).

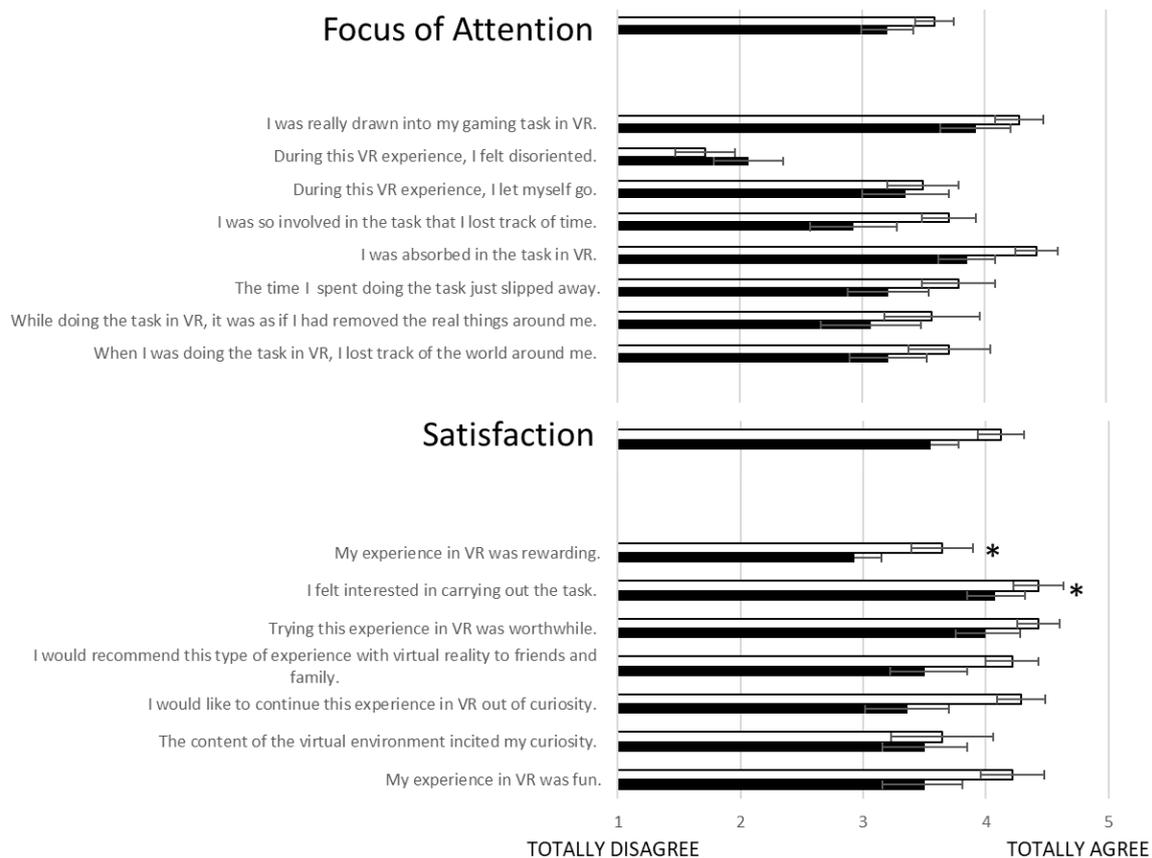


Figure 22 *STUDY 3 - VR EXPERIENCE*
Mean ratings for the Reaching (white bars) or Naming (black bars) group. Cumulative scores for focus of attention and satisfaction are highlighted in grey. The error bars represent the standard error of the mean.

4.5. Discussion

The cognitive ability to localize sounds relies on internal models that specify the correspondences between the auditory input reaching the ears, initial head-position and coordinates in external space. It is now well established that these correspondences can be updated throughout life (Carlile, 2014; Keating & King, 2015; Rabini et al., 2019; Strelnikov et al., 2011; Van Wanrooij & Van Opstal, 2005), setting the basis for re-learning spatial hearing abilities in adulthood. In addition, whenever auditory cues are altered, humans can exploit active perception strategies to adapt to the novel listening situations. For instance, people with hearing-impairment use head-movements to cope with their hearing difficulties (Brimijoin et al., 2010, 2012). Understanding the cognitive adjustments of internal models for spatial hearing, as well as the behavioral strategies adopted when localizing sounds in novel listening conditions, is particularly important for individuals who experience long-term auditory alterations (e.g., hearing loss, hearing aids, cochlear implants). In addition, it may provide useful indications for training normal hearing individuals who must adapt to novel auditory cues (e.g., when listening with non-individualized HRTFs in virtual auditory environments).

Until now, however, several methodological constraints have limited our understanding of the mechanisms involved in adapting spatial hearing skills when exposed to novel listening situations. Specifically, the potential role of dynamic and sensorimotor interactions with the sounds have remained largely overlooked. These interactions include head-movements performed during sound, as well as reaching actions towards the sounds themselves. Using a novel methodology, based on virtual reality and real-time kinematic tracking of the participant's head and hand, we examined the contribution of reaching movements and head movements when adapting to monaural listening. Our experimental approach is based on a new apparatus (Gaveau et al., 2020) that allows to present free-field sounds from a real loudspeaker, placed at pre-determined locations and aligned with a virtual reality scenario. This allows full control over visual cues about the visual environment (here, an empty room), cues to sound position (here, a visible array of virtual loudspeakers) and feedback (here, the flashing speaker that appeared in case of mistaken responses). Most importantly, our approach allows full monitoring of head position. This ensured identical straight-ahead initial posture for all trials and participants, but also permitted tracking of spontaneous head-movements during the response phase. Finally, precise kinematic of the hand response was measured in real-time. This allowed measuring of end-pointing positions in the Reaching group and ensured compliance with the instructions in the Naming group.

Four novel findings emerged. First, we documented that adaptation to monaural listening can occur over a relatively short period of time (about 200 trials, completed in approximately 50 min). Second, we showed that monaural plugging affected both sound localization performance and spontaneous head-movements. Third, in line with our main prediction, we found that reaching to sounds proved more effective than naming sound position when adapting to altered binaural cues. Fourth, we document that head-movements played an important role in this fast adaptation: reduction in localization errors was accompanied by wider portions of space explored with the head. Notably, performance improvements correlated with changes in head-movement extension selectively in the Reaching group, providing initial evidence for an interaction between reaching to sounds and head-movements when adapting spatial hearing abilities to monaural listening conditions.

4.5.1. Sound localization improves rapidly during monaural listening

The present study demonstrates that normal hearing adults, listening with one ear plugged and muffled, can improve their sound localization abilities over a short period of time. Both group of participants (Reaching and Naming) progressively reduced their error over trials, within a single testing session. We promoted adaptation to altered binaural cues through a combination of audio-visual feedback (Rabini et al., 2019; Strelnikov et al., 2011; Valzolgher, Campus, et al., 2020) and dynamic sensorimotor interactions with the sounds (spontaneous head-movements, agency over the sound). Despite most of the participants

were naïve to virtual reality, they easily understood the instructions, were engaged by the task, and gave a positive assessment to the VR experience overall. This confirms the feasibility of our novel approach, which represents an important pre-requisite for the use of similar VR applications in auditory perception research and clinical rehabilitation (see 3.4).

While adaptation to altered auditory cues has been documented in several previous reports (Carlile, 2014; Kacelnik, Nodal, Parsons, & King, 2006; Keating & King, 2015; Nawaz, McNeill, & Greenberg, 2014; Steadman et al., 2019a), to the best of our knowledge, improvements on a trial-by-trial basis have not been studied and described before. Traditionally, spatial hearing adaptation has been examined in paradigms which comprised pre- and post-testing phases separated by training sessions which occurred over several days (e.g., 5 days in Strelnikov et al., 2011 and in Irving and Moore, 2011; 10 days in Ohuchi et al., 2005). Thus, even when adaptation of spatial hearing abilities occurred after a small number of short training session (3 sessions, lasting 15 min as in Steadman et al., 2019), they were performed across multiple days. Although in our study we did not test the generalization of the observed adaptation of spatial hearing abilities, we investigated the improvements occurring over the course of 204 trials administered in a single session.

Two mechanisms may subtend the fast adaptation we have documented. On the one hand, it is possible that participants learned to adjust the contribution of the different auditory cues to sound localization. Reweighting of the auditory cues, with increased reliance on monaural cues at the unplugged ear for localization, is one of the mechanisms by which sound localization in azimuth can improve during monaural listening (Kumpik & King, 2019). We designed the experiment to selectively impact on binaural cues (both interaural-level differences and inter-aural time differences) while leaving monaural cues at the unplugged ear intact. Hence, participants may have learned to exploit monaural cues to a greater extent during the session. On the other hand, because participants were free to move their head during sound emission, it is possible that strategic (rather than cognitive) adaptation mechanisms could account for the observed fast improvements in sound localization. Cognitive and strategic contributions to sound localization improvements are clearly not mutually exclusive, and our study cannot determine their relative contributions to the observed performance changes. Yet, as described in the next section, we suggest a key role of compensatory behavioral strategies related to head-movements.

4.5.2. Monaural listening triggered compensatory head-movements

The role of head-movements in sound localization has been emphasized since the 1940's (Wallach, 1940), but it has been studied sporadically until recent years (Iwaya et al., 2003; Morikawa & Hirahara, 2013; Morikawa, Toyoda, & Hirahara, 2011; Perrett & Noble, 1997a). Here, we characterized head-movements with different variables (number and amplitude of head-movements, head-rotation around

vertical axis, and overall head-direction bias during an entire block). Our results show that head-movements changed immediately when passing from binaural to monaural listening conditions. In the first monaural listening block, the number of head movements increased, specifically for movements of smaller amplitude. In addition, participants turned their heads to approach sounds with their unplugged ear (as evidenced by immediate changes in the intercept). Our results also show that participants adapt their head-movements across the monaural blocks to adapt to the altered listening condition. Specifically, the number of head-movement continued to increase, but this time for movements of larger amplitude (compare Figure 19 B and A). Moreover, participants increased their exploration of auditory space by turning their head to more eccentric positions (as evidenced by changes in slope; see Figure 20 A) and turned their head even more to approach the sounds with their unplugged ear (see Figure 20 B).

Many previous works showed that head-orientation and head-movements improve sound localization performance in azimuth and elevation (Morikawa & Hirahara, 2013; Perrett & Noble, 1997b; Thurlow & Runge, 1967) and facilitate disambiguation of front/back confusions (Brimijoin & Akeroyd, 2012; Kim, Barnett-Cowan, & Macpherson, 2013; Perrett & Noble, 1997a; Pöntynen & Salminen, 2019; Wightman & Kistler, 1999a). This document the well-known effect of head-movements on spatial hearing. However, much less work has described the actual changes in head-movements that occur in difficult or altered listening conditions – as here. In a task that required head orienting to spoken sentences, Brimijoin et al. (2010) observed that normal hearing listeners performed smoother and more sigmoidal head-movement trajectories compared to hearing-impaired listeners. In the latter group, abrupt velocity changes, reversal of directions and frequent corrections tended to prevail (Brimijoin et al., 2010). Note however, that this study described the kinematic of head-movements when participants were explicitly instructed to point to the sound, whereas we reported the head-movement pattern observed in the spontaneous movements produced while the sound was playing and before the response. Although other works have examined spontaneous head-movements in novel and challenging listening conditions (as here), they did so when participants were not explicitly instructed to localize sounds. For instance, Brimijoin and Akeroyd (2012) has shown that participants use head-movements to increase the level of the signal in their better ear, or to increase the difference in level between the signal and the noise, when listening in complex auditory scenes (Brimijoin & Akeroyd, 2012). Remarkably, this type of head movement differs compared to those performed in more typical hearing situations (i.e., binaural hearing), whereby participants rather tend to face the speaker frontally when performing a sounds localization task (Iwaya et al., 2003; Thurlow & Runge, 1967). A recent study by Hadley and colleagues (Hadley et al., 2019) has shown that, as noise increases, people move closer to the target rather than make use of the potentially beneficial effects of head-orientations.

In our study, participants appear to have moved their head during monaural listening to achieve two goals: (1) orienting of their unplugged ear to the sound; (2) exploring auditory space to a greater extent. The need to cope with monaural listening was fulfilled by increasing the frequency and extent of head-turns to favor perception with the unplugged ear. This peculiar type of movement could have allowed greater exploitation of monaural cues. Specifically, listeners could have rapidly made use of changes in spectral monaural cues across the blocks, or – perhaps more likely – they could have exploited the attenuating effects of the head to identify the azimuthal position of sounds (for discussion of the different use of monaural cues in actual or simulated unilateral hearing loss see Kumpik and King, 2019). Along a similar line of reasoning, greater exploration of auditory space may have permitted participants to ‘search’ for the region of maximal sound intensity. A third possibility is that participants learned to use their asymmetric binaural cues to localize sounds in azimuth (as proposed by Agterberg et al., 2012, albeit for listeners with acquired unilateral conductive hearing loss who had much longer experience with the asymmetrical binaural cues) (Agterberg, Snik, Hol, Van Wanrooij, & Van Opstal, 2012). In particular, it is possible that they learned to determine the head-orientation that brought the target sound on their medial plane (i.e., 0 ITD orientation). In future works, it would be important to assess the relative contribution of spectral and intensity monaural information, by affecting the spectral information at the unplugged ear using an ear-mould.

4.5.3. Reaching to sounds promote adaptation to monaural listening more than naming

One clear result that emerged from our study is that the instructions on how to perform the task (Reaching vs. Naming) impacted on sound localization performance, head-movements during sounds and adaptation to monaural listening in general. In the first monaural listening condition (M1), the Naming group perceived sounds more toward the unplugged side, whereas the Reaching group did not present this bias (Figure 16 B). In addition, across trial repetitions, participants of the Reaching group reduced localization errors faster compared to those who interacted with sounds simply by reading labels (the Naming group). This result revealed that Reaching to sound was effective in promoting faster adaptation to monaural listening. Across monaural trials, the error changed differently. As indicated by the signed error (which refers to the lateralized bias of the response), the Naming group reduced the bias toward the unplugged side across trials, while the Reaching group developed a slight bias towards the plugged side (Figure 17 B).

Response type also affected head-movements during sound emission. Participants in the Reaching group turned their head as a function of target position to a greater extent compared to the Naming group, irrespective of listening condition (binaural vs. first monaural block; see 3.2.1). Interestingly, when

considering the progressive adaptation to monaural listening blocks, only the Reaching group showed a correlation between head-movement extension (slope) and performance (Figure 21). Thus, although participants in both groups were free to move the head during the task, only participants in the Reaching group appeared to have developed effective dynamic listening strategies involving head-movements, which in turn impacted on adaptation to monaural listening.

Why the response type (Reaching vs. Naming) influenced adaptation to monaural listening? As anticipated in the introduction, reaching to sounds requires a coordination of different effectors (eyes, head, hand) into a common reference frame (Cohen & Andersen, 2002). In turn, this may result in a more stable (or salient) spatial coding of sound source location and favor re-learning of sound-space correspondences. In addition, reaching movement may have helped directing attention towards the position occupied by the sound source, because of additional visual, proprioceptive and kinaesthetic cues resulting from the action. The analysis of spontaneous head-movements, however, revealed a more complex scenario than expected. Reaching participants explored a wider portion of space with the head compared to Naming participants, and this was functional for adaptation selectively for the Reaching group. Movements of smaller amplitudes, appeared to have played a less important role for adaptation: they were more numerous for Naming than Reaching participants, but only in the first monaural block. At present we can only speculate about the origin of this different head-movement behavior (and consequent change in performance) in the Reaching group. One possibility is that head-movements were more functional to the key task (sound localization) for the Reaching group than the Naming group. Specifically, it could be argued that participants in the Naming group used head-movements also to explore the labels that identified the speakers (recall that labels changed randomly between trials and thus they had to be read anew every time). In turn, this could have resulted in head-movements that could have interfered with the ones needed to optimally adapt to the monaural listening condition. Another possibility is that Reaching was a more engaging task for participants than Naming. In the questionnaires, the Reaching participants reported a stronger feeling of agency on the sound sources. In turn, this could have led to a stronger motivation when performing the task (Aarhus, Grönvall, Larsen, & Wollsen, 2011). The Reaching group also perceived the experience as more rewarding overall and expressed more interest in the task compared the Naming group. Recall that the participants in the Naming group did not directly interrupted the sound, as it was the experimenter that entered the response. Future works should make this aspect more comparable between groups and assess to what extent motivation and sense of agency per se could be at the origin of the different head-movement behavior.

4.6. Conclusions

In this article we show that reaching to sounds can promote fast adaptation to altered binaural cues, more than naming sounds. In addition, we document that listening with one ear plugged triggers compensatory head-movement behavior, which in turn can modulate sound localization performance. The head-movement behavior we observed can be interpreted as a strategic (rather than cognitive) adaptation mechanism to altered spatial hearing. However, cognitive and strategic contributions to sound localization improvements are clearly not mutually exclusive and, in fact, it is plausible that head-related behavioral strategies may trigger cognitive changes. The implications of these results are two-fold. First, they show the importance of considering dynamic listening behavior (reaching to sounds and head-movements) when studying how humans adapt to altered auditory cues. This is now increasingly possible thanks to virtual reality and kinematic tracking approaches, like the one we adopted in the present study and other recent works (Parseihian et al., 2014; Parseihian & Katz, 2012; Steadman et al., 2019a; Stitt et al., 2019). More generally, researchers should pursue the novel opportunity to study spatial hearing in more realistic and ecological environments (Hadley et al., 2019). Second, our findings could pave the way for new approaches to the rehabilitation of spatial hearing difficulties in people suffering from hearing loss or using hearing-enabling devices such as hearing aids or cochlear implants.

Chapter 5. Study 4: Adapting to altered auditory cues: generalization from manual reaching to head pointing

5.1. Abstract

The ability to localize sounds relies on processing of the auditory cues that result from interplay between the sound waves, the head and the ears. When the auditory cues change, sound localization becomes difficult and uncertain. The brain can adapt to altered auditory cues throughout life, and multisensory training can promote relearning of spatial hearing skills. Here, we study the training potentials of sound-oriented motor behavior, to test if a training based on manual actions toward sounds can produce learning effects that generalize to different auditory spatial tasks. Using visual virtual reality and body motion tracking, we assessed spatial hearing relearning in normal hearing adults with one-ear-plugged. Participants performed two auditory tasks that entail explicit and implicit processing of sound position (head-pointing sound localization and audio-visual cueing, respectively), before and after a spatial training session in which they identified sound position by reaching to near-by auditory sources. Using a cross-over design, we compared the effects of the spatial training with a control condition using the same stimuli, but different task demands (i.e., a non-spatial discrimination of amplitude modulations in the sound). Our findings show that reaching to sounds improves spatial hearing in participants using an ear-plug, more than the control condition. Training by reaching also modified head-movement behavior. Crucially, the improvements observed during training generalize also to a different type of sound localization task, possibly as a consequence of acquired and novel head-movement strategies.

5.2. Introduction

Humans and other hearing species can localize sounds in space. Spatial hearing relies on the interpretation of binaural auditory cues, resulting from the difference in the inputs reaching the two ears, and monaural cues, resulting from the amplitude and spectral changes occurring at the single ear (Middlebrooks, 2015; Middlebrooks & Green, 1991; Risoud et al., 2018). Although listeners experience their spatial hearing skills as constant and stable, temporary or permanent conditions can alter the auditory cues and affect this fundamental hearing ability. Examples include partial or complete hearing loss to one ear (e.g., Kumpik & King, 2019), age-related hearing loss (e.g., Dobreva, O'Neill, & Paige, 2011), use of hearing aids (e.g., Van den Bogaert, Klasen, Moonen, Van Deun, & Wouters, 2006) or use of cochlear implants (e.g. Akeroyd, 2014). Yet, research in the last decades has shown that relearning of spatial hearing skills is possible, in humans (Carlile, 2014; Keating & King, 2015; Mendonça, 2014; Rabini et al., 2019; Strelnikov et al., 2011; Valzolgher, Campus, et al., 2020) as well as other animals (Kacelnik, Nodal, Parsons, & King, 2006).

Adaptation to new auditory cues can be achieved using multisensory cues to sound position (for reviews see: Carlile, 2014; Irving & Moore, 2011; Keating & King, 2015; Knudsen & Knudsen, 1985; Mendonça, 2014). Training with audio-visual stimuli can be more effective than training with auditory information alone

(Majdak, Walder, & Laback, 2013; Rabini et al., 2019; Strelnikov et al., 2011; Zahorik, Bangayan, Sundareswaran, Wang, & Tam, 2006). In addition, adaptation to new auditory cues may be promoted through motor interactions with the sound sources, using tasks in which the acoustic stimulation results from the subjects' own movements (e.g., Valzolgher, Campus, et al., 2020) or in which participants are encouraged to act towards the sounds (e.g., Parseihian & Katz, 2012; see also Mendonça, Campos, Dias, & Santos, 2013). For instance, participants could be asked to hit a moving sound presented in virtual auditory space using a hand-held tool (Honda et al., 2007; Ohuchi et al., 2005), or to shoot audio-visual moving targets presented again in virtual reality (Poirier-Quinot & Katz, 2020). Other relevant examples include the work of Steadman and colleagues, in which participants were required to point actively the head toward sounds (Steadman, Kim, Lestang, Goodman, & Picinali, 2019). Taken together these works suggest that motor interactions with sounds could promote adaptation to new auditory cues.

In one recent study (Valzolgher, Verdelet, et al., 2020), we examined this issue directly by asking normal hearing participants with one ear plugged to identify the position of sounds by reaching or by naming the labels associated to each speaker. The manipulation was performed across groups, using virtual reality (VR) to present 17 virtual speakers in front of the participant in a virtual room and kinematic tracking to monitor hand movement and head rotations. Both groups received audio-visual feedback about their performance on a trial-by-trial basis. Importantly, participants in either group were always allowed to move their head during each sound, which lasted until response (approximately 4 seconds). Results showed that the reaching group decreased sound localization errors more rapidly compared to the naming group. Moreover, the reaching group increased head exploration movements during listening, and these head-movements were correlated with localization improvement rate. This suggests a specific role for reaching to sounds when adapting to new auditory cues. In addition, it points to potential role head movements in this adaptive behavior.

Valzolgher et al. (2020) documented that reaching to sounds can ameliorate sound localization with one ear plugged on a trial-by-trial basis. However, it remained unclear if participants improved because of practice with the specific auditory task or instead they learned new and effective ways to adapt to the altered auditory cues. If the latter is true, then sensorimotor training should improve sound source localization also when the task entails different sound source positions, and when it requires a response with a non-trained effector (i.e. the head rather than the hand). In the present study, we directly addressed this hypothesis by testing if performance improvement induced by reaching to sounds can extend (i.e., generalize) beyond the trained auditory task itself.

To this aim, we recruited a group of normal hearing participants and we temporarily degraded their spatial hearing by plugging one ear. We used the reaching task developed in our previous work (Valzolgher, Verdelet, et al., 2020) to train their sound localization performance. Crucially, before and after this training

we also tested participants in two tasks aimed at revealing generalization effects: a head-pointing sound localization task (Valzolgher, Verdelet, et al., 2020) and an audio-visual cueing task (Koelewijn, Bronkhorst, & Theeuwes, 2009; Spence & Driver, 1996). The head-pointing localization task required to explicitly localize sounds and differed with respect to training in terms of spatial position of the targets and with respect to response effector (the head instead of the hand). The rationale was to minimize any potential effect of sensorimotor adaptation. The audio-visual cueing task was instead an implicit sound localization task, in which sounds serve as lateralized attention-orienting cues for the discrimination of visual targets (i.e., an audio-visual analogue to the classic attention cueing paradigm) (Posner, 1980; Spence & Driver, 1996). This second test was introduced to probe whether adaptation to new auditory cues could impact also on audio-visual attention orienting, a skill that can be hampered by monaural listening (Pavani et al., 2017).

Using a crossover experimental design (Figure 23), we compared a training in which participants performed reaching movement to identify sound position (hereafter referred to as Spatial training) with a control training in which participants performed comparable reaching movements to identify a non-spatial feature of the sound (from now on, the Non-Spatial training). In the Non-Spatial training participants were invited to discriminate between sounds with two different amplitude modulations rather than focusing on the spatial position of the sources. Each participant received both training conditions, in two successive sessions separated by a wash-out period (2-weeks minimum). While the two trainings used identical stimuli, the behavioral requests emphasized Spatial processing vs. Non-Spatial processing of sounds. These different task demands recruit substantially different cognitive and brain mechanisms, as revealed by converging evidence from neurophysiology (Rauschecker, 2018), neuropsychology (Clarke, Bellmann, Meuli, Assal, & Steck, 2000) and cognitive neuroscience (Ahveninen et al., 2006; Alain, Arnott, Hevenor, Graham, & Grady, 2001; Noyce, Cestero, Michalka, Shinn-Cunningham, & Somers, 2017). Our key prediction is to observe generalization of training effects after the Spatial training more than after the Non-Spatial training paradigm.

5.3. Methods

5.3.1. Participants

Twenty participants (age: $M = 29.4$, $SD = 10.5$, 5 males, 19 right-handed) were recruited to participate in the study, which was carried out in otolaryngology department of the civil hospital Edouard Herriot (HEH) in Lyon (France). All participants signed an informed consent before starting the experiment, which received ethical approval from national ethics committee (Ile de France X, N° ID RCB 2019-A02293-54), and recorded in clinicaltrials.gov (NCT04183348).

All had normal to corrected-to-normal vision and reported no motor or vestibular deficit as well as no history of neurological or psychiatric disorders. Hearing thresholds were measured for all participants using an audiometer (Equinox 2.0, Interacoustics), testing different frequencies (250, 500, 1000, 2000, 4000, 8000 Hz), on the right and left ear separately. All participants had an average threshold for both ears below 20 dB HL.

5.3.2. General structure of the experimental session

Participants were invited to participate in two experimental sessions, separated by at least 2 weeks. Each session comprised three phases: two testing phases and one training (Figure 23 A). The two experimental sessions differed only in the task required during the training phase: participants were involved once in the experimental training task (Spatial Training) and the other time in the control training task (Non-Spatial Training). In this way, all participants performed both training type in a within-subject design. Order of training type was counterbalanced across participants. The testing phases were identical in both experimental sessions: they included the head-pointing sound localization (conducted in VR) and the audio-visual cueing task (conducted outside VR).

Participants completed the entire experiment in monaural listening. This temporary auditory cue alteration was obtained by occluding the right ear of the participant with a plug (3M PP 01 002; attenuation values: high frequencies = 30 dB SPL; medium frequencies = 24 dB SPL; low frequencies = 22 dB SPL) and a monaural ear muffs (3M 1445 modified to cover only the right ear; attenuation values: high frequencies = 32 dB SPL; medium frequencies = 29 dB SPL; low frequencies = 23 dB SPL). At the beginning of their first session, participants the testing phase was performed also without the ear plug (Figure 23 A). This provided a baseline measure of their spatial hearing skills in binaural listening before being exposed to the monaural listening condition.

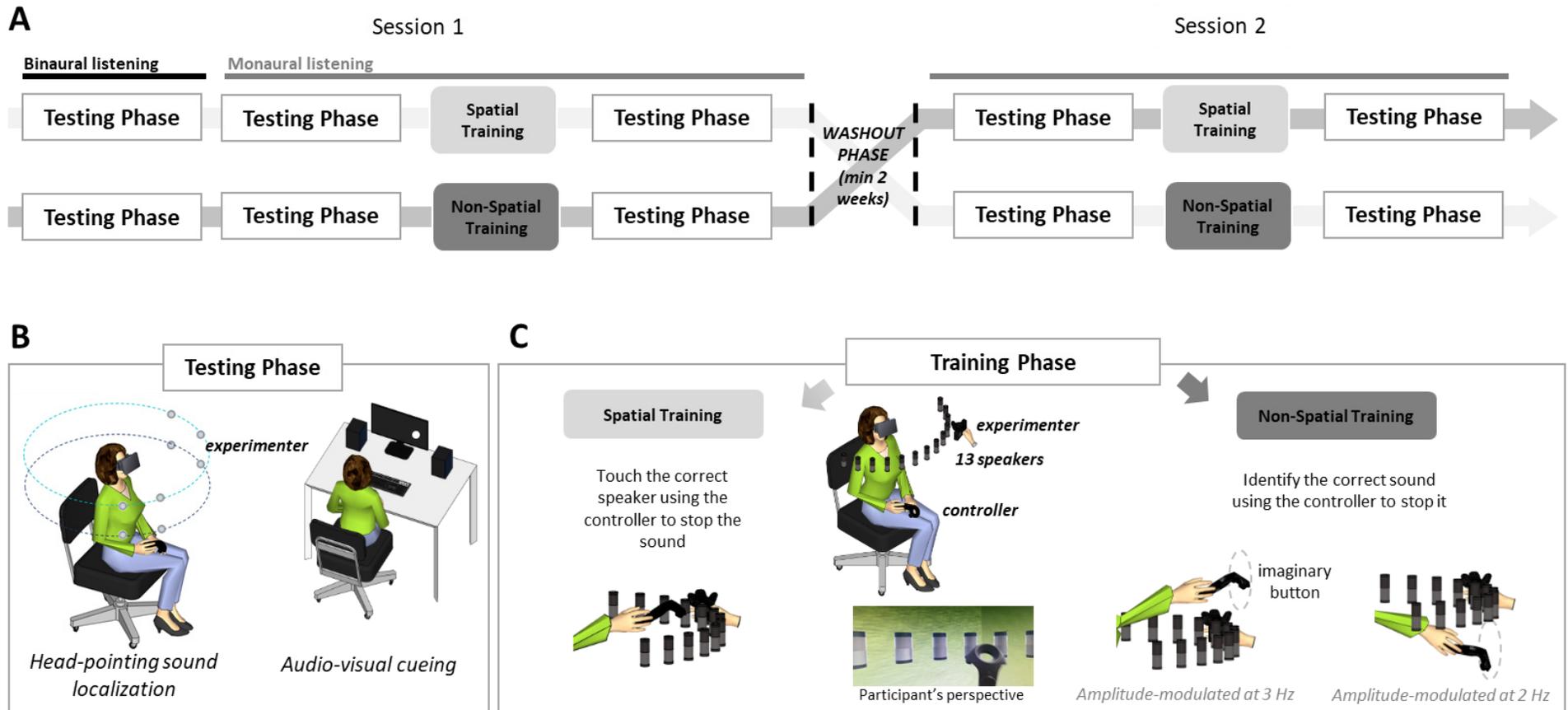


Figure 23 STUDY 4 - EXPERIMENTAL PROCEDURE AND SETTING

(A) Experimental cross-over protocol: all participants took part in 2 training session (session 1 and session 2). Each session was composed by two Testing phases, completed before and after Training phase (Spatial or Non-Spatial). All phases were performed in monaural listening (i.e. with one ear plugged), but an additional testing phase at the beginning of session 1 measured performance in binaural listening. (B) Testing Phases. To the left, a schematic representation of participant wearing the HMD and holding the controller used for validation during head-pointing sound localization. The eight spheres in front of the participant indicate the pre-determined speaker positions (see text for details). To the right, a schematic representation of the setting for the audio-visual cueing task. (C) Training Phase. In the center, a schematic representation of participant wearing the HMD and holding the controller with the 13 cylindrical visible speakers indicating the pre-determined speaker positions used in the two-training types. Below, a close-up of the scene as visible inside the HMD from participant's perspective: during the task participant saw the virtual room, the visible speakers and the controller they were holding in their hand. On both sides, the response's movements during both the Spatial and Non-Spatial training.

5.3.3. Apparatus

Virtual reality and kinematic tracking were implemented using the HTC Vive (Vive Enterprise). The system comprised one head-mounted display (HMD, resolution: 1080 x 1200 px, Field Of View (FOV): 110°, Refresh rate : 90 Hz), 2 hand-held controllers (one held by the experimenter to calibrate head center, and one held by participants to interact with the virtual environment during the training phase), 1 tracker mounted above the speaker to track its position in real time, and 2 lighthouse base-stations (Lighthouse V1.0) for scanning the position of the controller, trackers and the HMD. Tracking precision and accuracy of the HTC Vive System is adequate for behavioral research purposes (Verdelet et al., 2020). Stimuli were controlled and delivered using a LDLC ZALMAN PC (OS: Windows 10 (10.0.0) 64bit; Graphic card: NVIDIA GeForce GTX 1060 6GB; Processor: Intel Core i7-7700K, Quad-Core 4.2 GHz/4.5 GHz Turbo - Cache 8 Mo - TDP 95W) using Steam VR software and the development platform Unity3D (Unity Technologies, San Francisco, CA).

Real free-field auditory stimuli were delivered by a loudspeaker (JBL GO Portable from Harman International Industries, Northridge, California USA, 68.3 x 82.7 x 30.8 mm, Output Power 3.0W, Frequency response 180 Hz – 20 kHz, Signal-to-noise ration > 80 dB), with the HTC Vive tracker firmly attached to its top. During the entire experiment we tracked the loudspeaker position, as well the position of the controller held in the participant's hand and the HMD, via a dead reckoning process using gyroscope and accelerometer, plus a correction signal from the lighthouse system every 8.333 milliseconds. Both tracking method allowed position tracking with a frequency sample of 250 Hz. The software is designed to guide the experimenter to align the real loudspeaker (i.e., the sound source) with a set of pre-determined positions defined in the 3D virtual environment in each trial. This method combining virtual reality and kinematic tracking to measure sound localization abilities has been developed in our laboratory and has been already adopted in previous studies (Valzolgher et al., 2020).

The Audio-Visual Cueing test was carried out without VR, with the participant sitting at a desk. The apparatus for this task included a separate PC, a DELL 24" monitor, a keyboard and two speakers, positioned at ear level on both sides of the screen, each located 20° to the left or to the right of central fixation (see Figure 23 B). The height of the chair on which the participants sat was adjusted to favor the support of the head on the chin rest on the edge of the table, aligned with the center of the monitor. The test consists of a visual discrimination task implemented with the program OpenSesame®.

5.3.4. Procedure and stimuli

Before starting the experiment, participants were informed about the use of the VR equipment. When engaged in the VR tasks participants sat on a rotating armless chair with no chin rest, placed in the center of the room. The room (3.6 m x 3.9 m e 2.7 m) was quiet and treated with sound-proof panels to partially reduce echoes. The

T60 reverberation of the room was 0,30-0,33 seconds, as measured using a Blue Solo 01bB phonometer. Instruction for each task were provided immediately before the task started.

5.3.5. 2.4.1. Testing Phase

Head-pointing sound localization. This part of the experiment was carried out entirely in VR. The participant was immersed in a virtual room with green walls, reproducing exactly the size and shape of the real room. The room was devoid of any objects, except for light source on the ceiling and a wooden door behind the subject. The task comprised 40 actual trials, plus 5 practice trials presented at the beginning of the task. At the beginning of each trial, participants were asked to direct their gaze in front of them to align their head with a white central fixation cross. As soon as the head correct position was reached, the fixation cross turned blue. This procedure ensured that initial posture was comparable across trials and participants, even in the absence of a chin-rest. In the meanwhile, the experimenter placed the speaker at one of the possible eight pre-determined positions. Eight predetermined positions were used throughout the experiment, resulting from the combination of 4 different azimuths in the frontal space in respect to participant's head position (-67.5° , -22.5° , 22.5° or 67.5°), 2 different elevation (+ and -10°) and a single distance (55 cm) (Figure 23 A). Each position was reached manually by the experimenter using visual indications provided on the dedicated instruction monitor. When the correct starting head posture was reached, and the loudspeaker was positioned correctly (i.e., within a sphere with diameter of 5 cm centered on the pre-determined location), the target sound was delivered.

The target sound consisted of 3 second white noise bursts, amplitude-modulated at 2.5 Hz, and delivered at about 65 dB SPL, as measured from the participant's head. During target sound delivery the fixation cross turned white. From this moment, participants were free to move their heads and rotate the chair they sat on to explore the surrounding space. The task consisted in localizing the exact source of the sound and to indicate it using the head direction as a pointer. This response was aided by the fact that the visible fixation cross in the HMD followed head direction displacements. It is noteworthy that, since the speaker was invisible in VR, participants did not have visual cues about sound source position. At the end of the 3 seconds of sound presentation, the central cross turned red, to indicate to the participants to validate their response (i.e., their current head direction) by pressing the button at the base of their hand-held controller.

Participants were informed that sounds could be delivered anywhere in the 3D around them and they did not have to judge the distance but only the elevation and the azimuth dimension of the sound space. Note that head and trunk movements remained unconstrained both during and after sound emission, allowing spontaneous active listening behavior (e.g., orienting the head in the direction of the sound even during the sound emission).

Audio-Visual Cueing task. This part of the experiment was carried out entirely outside VR. Participants sat at the experimental table, placed inside the same room in which the VR experiments are carried out. Unlike all the VR

tasks, in this audio-visual cueing task participants rested their heads on the chinrest, hence no head-movements were ever allowed.

Each trial started with a white fixation cross appearing in the center of the monitor and remaining visible until response. After a random delay (450–600 ms), an auditory stimulus (white noise, duration 100 ms) was emitted from one of the two loudspeakers. Loudness of the auditory stimulus was approximately 60 dB SPL, as measured from head position. After 100 ms from sound delivery, the visual target was presented. This consisted of a white filled circle (20 pixels radius, 0.5° of visual angle), appearing on a black background for 140 ms in the upper or lower hemifield with respect to the horizontal meridian passing through visual fixation (1.15° above or below the meridian), either in the left or right hemifield (10° from fixation). In half of the trials, the visual target and the sound appeared on the same hemisphere (congruent trials), whereas on the remaining half the visual target and the sound appeared on opposite hemispaces (incongruent trials).

Participants were asked to keep their gaze toward central fixation throughout the task and to indicate as quickly and accurately as possible the elevation of visual targets appearing one at a time on either side of fixation. Up/down responses were given using the up/down arrows keys on an external keyboard, using the index-middle fingers of the right hand. Participants had a timeout of 2000 ms to give their answer. The experiment started with 8 practice trials, following by 128 trials divided in 2 blocks randomly divided in congruent and incongruent audio-visual conditions and it lasted 10 minutes approximatively (Pavani et al., 2017). Participants received feedback on accuracy (percentage of correct responses) and mean response time (in ms) only at the end of each block. Importantly, they were also explicitly informed that sounds were always entirely task-irrelevant.

5.3.6. Training Phase

This part of the experiment was carried out entirely in VR. In each session, this phase was either a Spatial or a Non-Spatial training (see below). Both training tasks took place in the same virtual room used for the head-pointing sound localization task. The only difference was that during both training tasks, thirteen virtual speakers were visible in front of participants. They were arranged in a semicircle, at the ear level, and distributed in front of participants spanning between $\pm 72^\circ$ (12° between each of them). The distance to the participant was always 55 cm (Figure 23 C). Note that the sound positions were different as compared to head-pointing sound localization.

Target sounds were delivered from the same azimuth and elevation as the virtual speakers, with a small distance offset so that the real speaker was 5 cm further away with respect to the virtual speaker to avoid collisions between the controller and the real speaker during the reaching response (see below). The target stimulus always consisted of a white noise: half of the stimuli were amplitude-modulated at 2 Hz and the remaining half at 3 Hz, to create clearly distinguishable targets. Targets sounds were delivered from each of the virtual speaker

in random order (12 repetition for each of the 13 speakers, resulting in 156 trials overall, divided into 3 blocks of 52 trials each).

Crucially, exactly the same stimuli were delivered in the two training tasks, thus making the auditory component of the two trainings identical in all respects. In addition, they both involved a similar motor response: a reaching movement. Participants were informed that at the beginning of each trial the controller had to be held in the right hand, at sternum level, and that once the response was completed it had to return to that starting position. However, for the Spatial training the reaching movements served to indicate the perceived sound position, whereas for the Non-Spatial training they served to indicate the perceived amplitude modulation in the target sound.

Spatial Training. Participants had to identify the speaker from which the sound was coming by reaching it using the controller in the right hand (Figure 23 C). To prevent participants from colliding with the speaker held in the experimenter's hand, the setting provides that the real speakers was moved back 5 centimeters in depth, which does not affect the directionality of the perceived sound. Participants received a brief vibration of the controller upon contact with the chosen speaker, irrespective of whether the response was correct or incorrect. Importantly, if they reached and touched the correct speaker, the sound stopped. On the contrary, if participants reached the wrong speaker they received a visual feedback: the correct speaker started to flash. Specifically, from the correct source location, a series of red concentric circles (1024x1024 px, 2 circles per second) expanded intermittently, irradiating the surrounding space. The rationale was to capture participants attention even if they were looking toward a different zone of the space, including the opposite hemispace. The visual feedback and the sound stopped only when the subject reached the correct speaker position with the controller. This has two implications: first, a sense of agency was associated with the correct response; second, whenever the wrong speaker was originally selected, a combination of visual and auditory signals guided the participant to the correct sound source.

Non-Spatial Training. Participants had to identify whether the sound emitted was amplitude modulated faster (3 Hz) or slower (2 Hz) rate (note that participants accustomed with the two auditory stimuli before starting the training). For the fast amplitude-modulated sounds, participants directed the controller in front of them, above the row of speakers arranged at head height, aiming to touch an invisible virtual button above the central speaker. For the slow amplitude-modulated sounds, participants had to reach an invisible virtual button below the same central speaker (Figure 23 C). A vibration from the controller indicated that one of two buttons was correctly reached. As in the Spatial training feedback procedure, a visual feedback was delivered in case of erroneous responses. This was a series of red concentric circles that expanded intermittently from above or below the central speaker to indicate the correct position to reach. Recall that target sounds were nonetheless

presented from different spatial positions during the training, although this spatial information was now totally task-irrelevant.

5.3.7. Analysis

Statistical analyses were run using R (version 1.0.143) (R Core Team, 2013). For the linear mixed-effect (LME) model, we used the R-packages *emmean*, *lme4*, *lmerTest* in R Studio (Bates, Mächler, Bolker, & Walker, 2015; Fox & Weisberg, 2020). The R-package *car* was used to obtain deviance tables from the LME models.

5.3.7.1. Performance

Head-pointing sound localization. To study performance in the head-pointing sound localization task, we focused on absolute error (i.e., the absolute deviation of the mean response from the source position) and signed error (i.e., the signed difference between the source and the response) in azimuth and elevation, separately. Signed error was negative or positive, to indicate an overall bias to respond. All variables were calculated for each individual trial. Trials in which the participant had problems with response validation (e.g., they responded before the end of sound emission), or in which the HMD signal were lost were removed from the analyses.

Audio-Visual Cueing task. To study performance in the Audio-Visual Cueing task we calculated the Cueing Effect (CE), expressed in milliseconds. The CE was calculated by subtracting the average reaction time (RTs) when the auditory cue and the visual target occur on the same side of space, from that obtained when the auditory cue and the visual target occur on opposite sides of space (Spence & Driver, 1996). The calculation was performed separately for each participant, phase and training. Trials with incorrect responses (i.e., wrong elevation judgement on the visual target) were excluded from the analyses.

Training task. To study changes in sound localization performance during the Spatial Training, we examined the absolute error (degrees) and the signed error (degrees) during the reaching to sound task. Because target position changed only in the horizontal dimension, all errors were computed only in azimuth. To study performance during the Non-Spatial training, we considered the percentage of correct answers.

5.3.7.2. Spontaneous head-movements

We measured head movement in all VR tasks (i.e., head-pointing to sounds, reaching to sounds during the Spatial Training, reaching to indicate amplitude-modulation differences in the Non-Spatial training).

To study head-movements, we calculated the tangential velocity of the head on the x, y, z axis (expressed in degrees of rotation) using two-points central difference derivate algorithm (Bahill & McDonald, 1983) with 5 points for the half-window (smoothing purpose). The onset and the end of the movements were computed automatically using a velocity-based threshold (10°/s) (as in Valzolgher, Verdelet, et al., 2020). Each head-

movement was checked manually by visualizing the spatial rotation changes of the head and its speed using a custom-made tool box in MATLAB R2018b. The rationale was to eliminate trials in case the HMD data were lost.

We focused the analysis on three dependent variables: the number of head-movements, the head-rotation extent around the vertical axis and the head-rotation bias (i.e., center of gravity of head-rotation). To compute number of head-movements, we considered all the detected peaks of velocity in the head trace, but we removed the movements smaller than 2° degrees to avoid noise (i.e. excluding movements which are not indicators of spontaneous head intentional movements and may reflect micro postural movements not related to the task). To calculate head-rotation extent we sum the absolute value of the rightward and leftward head-rotation extremity. For instance, if the head rotated 20° to the right and 40° to the left, the head-rotation extent was calculated as the sum of the two, hence 60°. To calculate head-rotation bias, we sum values of the rightward and leftward head-rotation extremity. For instance, if the head rotated 20° to the right and 40° to the left (left is expressed with negative sign: -40°), the head-rotation bias was calculated as the signed sum of the two, therefore -20°.

5.4. Results

5.4.1. Does monaural listening increase sound localization errors?

We started by validating the efficacy of our ear plugging procedure in reducing sound localization abilities. Figure 24 A shows the distribution of head-pointing responses in azimuth at the beginning of the experiment (i.e., session 1, before training), separately for binaural and monaural listening conditions. As expected, plugging the right ear increase uncertainty in sound localization responses and created a leftward bias in head-pointing, as evidenced by wider distributions and distribution peaks to the left of actual speaker locations (dashed lines).

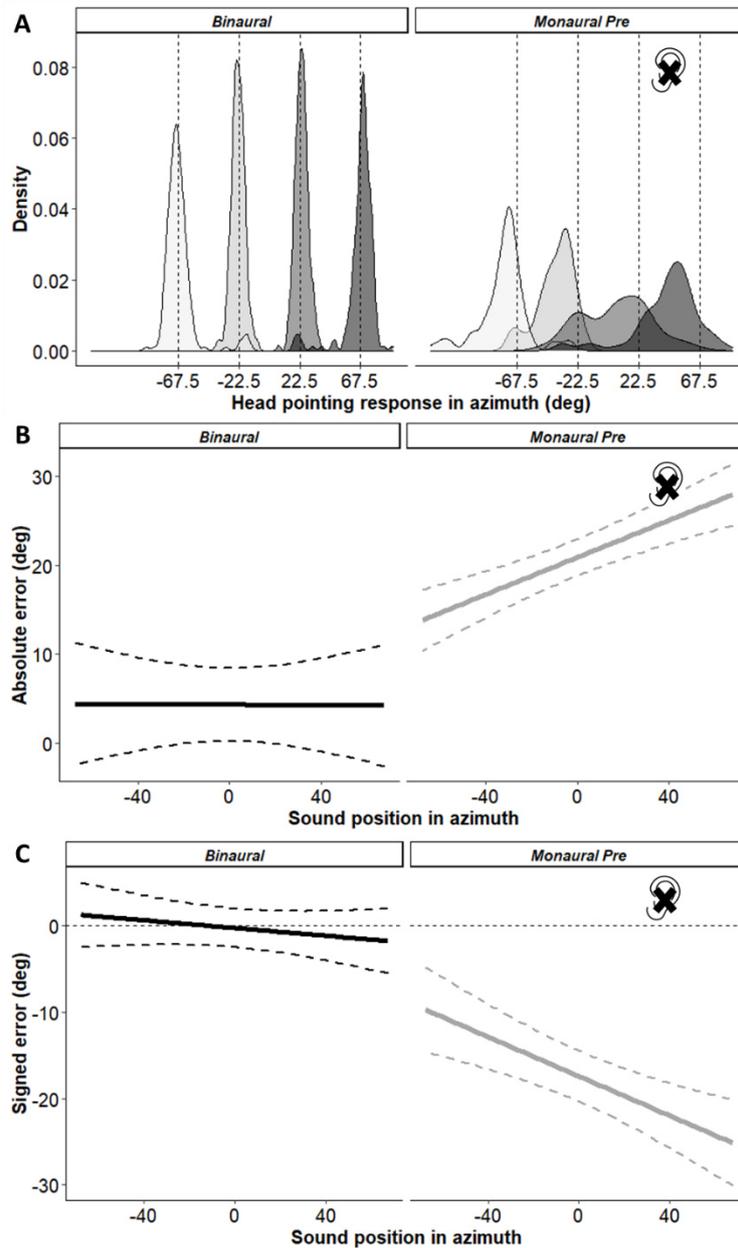


Figure 24 STUDY 4 - HEAD-POINTING SOUND LOCALIZATION IN SESSION 1 BEFORE TRAINING

(A) Density plot of responses in azimuth (in degrees), as a function of listening condition (binaural vs. monaural pre-training). (B) absolute error and (C) signed error, as a function of sound position in azimuth and listening condition. Dashed lines indicate the 95% confidence intervals of the overall linear regression.

To confirm the immediate effects of monaural plugging before training, we studied absolute and signed errors using separate linear mixed-effects (LME) models with LISTENING (binaural or monaural) and AZIMUTH as fixed effects, and PARTICIPANT (intercept and slope) as random effect. We found the main effect of LISTENING (absolute error: $X^2(1) = 882.70$, $p < 0.001$; signed error: $X^2(1) = 785.57$, $p < 0.001$) and the interaction between LISTENING and AZIMUTH (absolute error: $X^2(1) = 91.26$, $p < 0.001$; signed error: $X^2(1) = 57.39$, $p < 0.001$). Compared to binaural listening, errors increased after ear-plugging (mean \pm SD; absolute error: binaural = $4.3^\circ \pm 6.5^\circ$, monaural = $20.9^\circ \pm 16.0^\circ$; signed error: binaural = $-0.2^\circ \pm 7.6^\circ$, monaural = $-17.4^\circ \pm 19.2^\circ$). This was particularly

evident for targets delivered more toward the plugged side (see plots of absolute and signed error in Figure 24 B and 24 C).

Monaural plugging affected sound localization also in elevation (Figure 25). We entered absolute and signed errors in elevation in separate LME analyses with LISTENING and ELEVATION as fixed effects, and PARTICIPANT (intercept) as random effects. We found a main effect of LISTENING (absolute error: $X^2(1) = 13.33, p < 0.001$; signed error: $X^2(1) = 104.6, p < 0.001$). Localization errors in elevation changed after ear-plugging compared to binaural listening (mean \pm SD; absolute error: binaural = 10.6 ± 6.7 , monaural = 9.4 ± 4.4 ; signed error: binaural = 9.8 ± 7.5 , monaural = 5.3 ± 6.3).

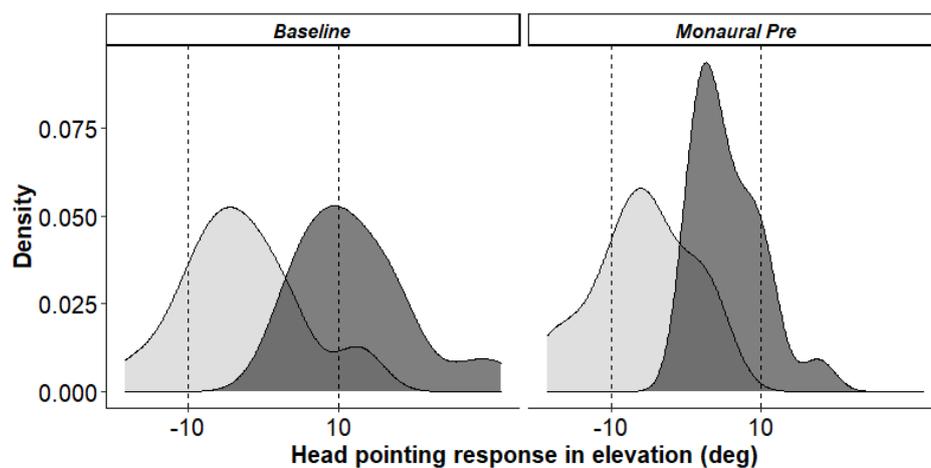


Figure 25 STUDY 4 - HEAD-POINTING SOUND LOCALIZATION, SESSION 1
Density plot of response in elevation as a function of sound position in elevation (-10° or 10°).

5.4.2. Does spatial training reduce sound localization errors?

Next, we turned to examine if the spatial training was effective in reducing the sound localization error caused by monaural listening. Specifically, we studied if participants adapted to monaural listening across successive trials. To this aim, we entered absolute and signed errors into separate LME models with TRIAL NUMBER as fixed effect (as in Valzolgher et al., 2020). We also included in the model PARTICIPANT (intercept and slope) and SESSION (intercept) as random effects. The latter was added to account for the variability related to the session in which the Spatial training was completed (first or second). The analysis on absolute error revealed a main effect of TRIAL NUMBER, $X^2(1) = 4.11, p = 0.04$. As shown in Figure 26 A, the absolute error reduced across trials. The analysis on signed error revealed no main effect of TRIAL NUMBER emerged ($X^2(1) = 0.77, p = 0.38$), but it is noteworthy that that leftward bias in sound localization decreased numerically, approaching zero (see Figure 26 B)

The Non-spatial task was completed with great accuracy by all participants (mean accuracy = 99.04%; SD = 1.09%) already from trial 1 and throughout the training session (Figure 26 C, lower panel).

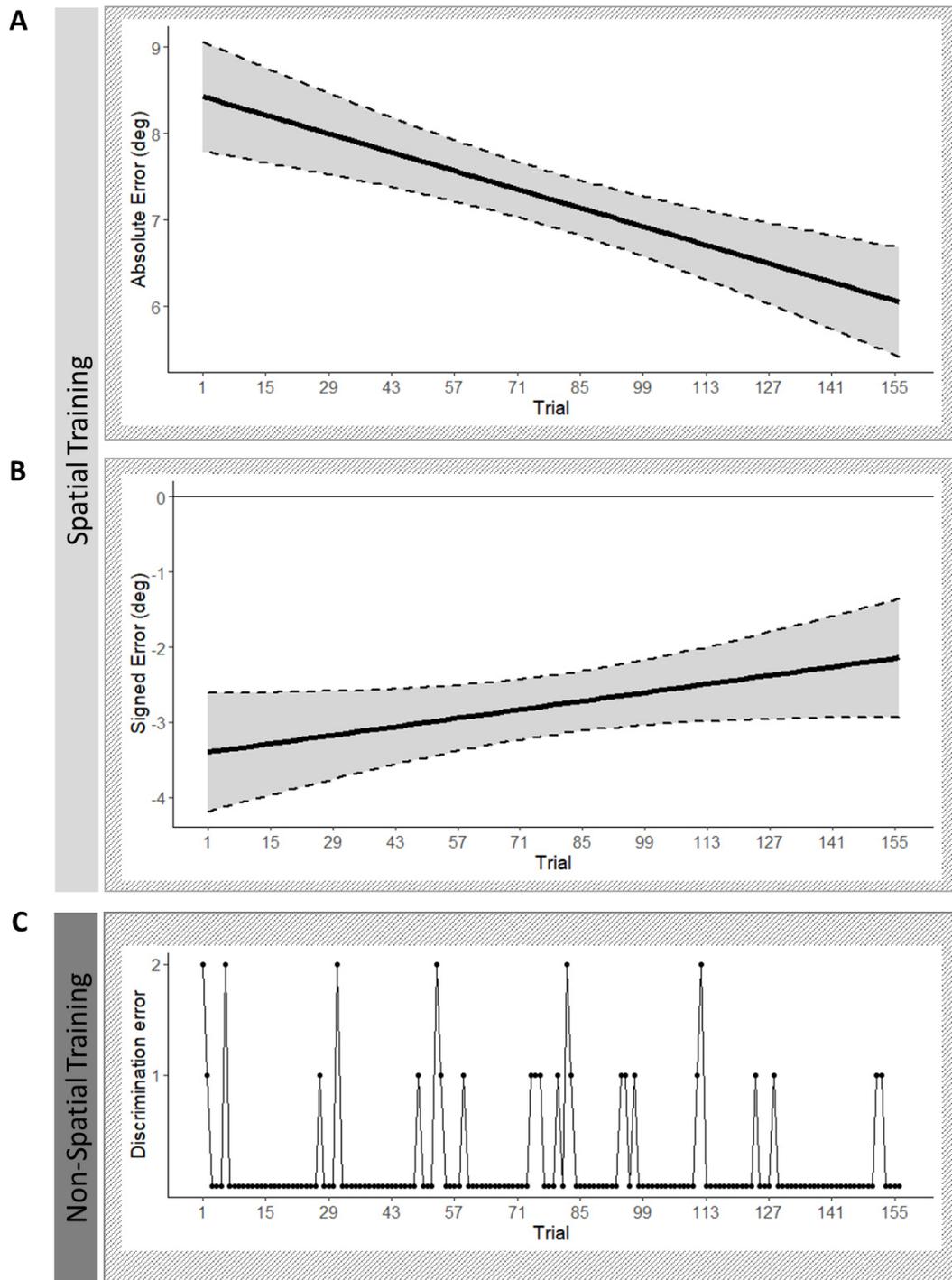


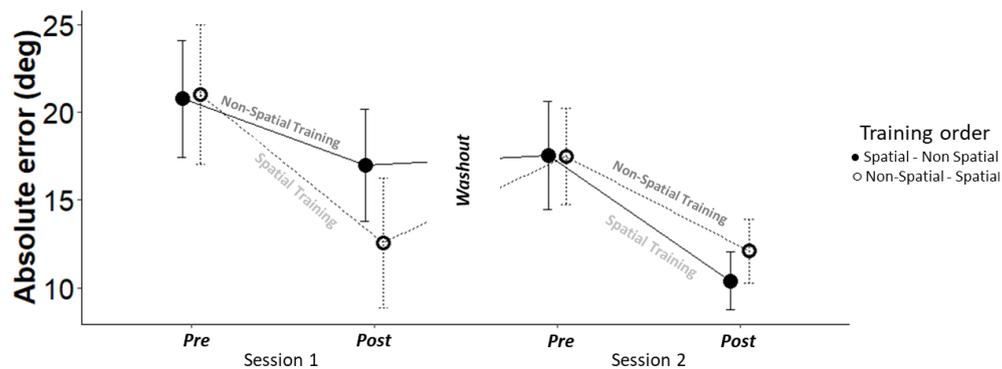
Figure 26 *STUDY 4 - PERFORMANCE DURING SPATIAL AND NON-SPATIAL TRAINING*
 Reduction of absolute (A) and signed (B) error as a function of trial in the Spatial training. (C) Cumulative discrimination error across participants (i.e., number of participants who made a mistake in the trial), shown as a function of trials in the Non-Spatial training. Negative values of signed error indicate a bias toward the unplugged side. Dashed lines indicate the 95% confidence intervals of the overall linear regression.

5.4.3. Does Spatial training effects generalize to the head-pointing sound localization task?

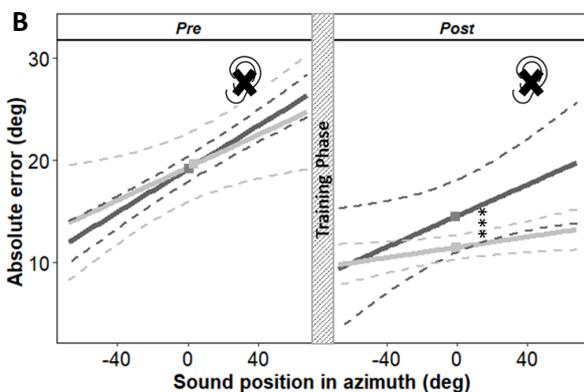
Having documented that the spatial training improved sound localization, we tested our key hypothesis about generalization of training effects to other sound localization tasks. Figure 27 A shows the progression of absolute

localization errors in head-pointing sound localization across the two sessions of the experiment, separately for the testing sessions before training (Pre) and after training (Post). Participants who underwent the Spatial training on session 1 and those who underwent the same Spatial training on session 2 (i.e., started with the Non-spatial training instead) are indicated by separate lines (dashed vs. continuous line, respectively). Three aspects are clearly visible in Figure 27 A: (1) both trainings improved performance; (2) the Spatial training improved performance to a greater extent compared to the Non-Spatial training; (3) the interval between the two testing sessions (session 1 and session 2) was effective in inducing a partial wash-out of training effects, making the two groups again comparable in the pre-training session of session 2.

A



B



C

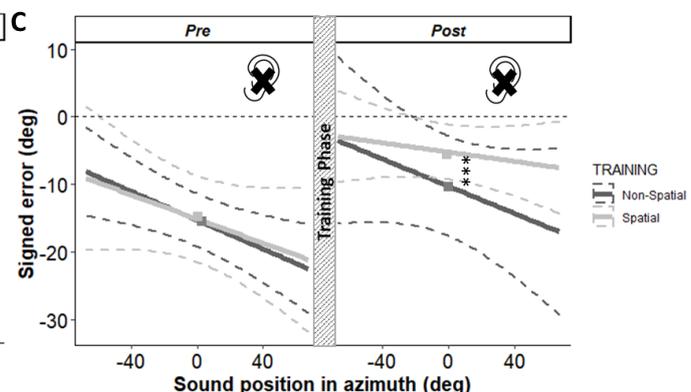


Figure 27 STUDY 4 - PERFORMANCE IN HEAD-POINTING SOUND LOCALIZATION

Top: (A) Progression of absolute localization across the four testing sessions of the experimental design, separately for participants who completed the Spatial training on session 1 (dashed line) or session 2 (continuous line). Bottom: Absolute (B) and signed (C) errors (in degrees) in monaural listening are plotted as a function of sound position in azimuth, separately for each training type (light grey: Spatial training; dark grey: Non-spatial training). For the signed error negative values indicate a bias toward the unplugged side. Dashed lines indicate the 95% confidence intervals of the overall linear regression.

To directly compare the effects of training type on head-pointing sound localization, we entered absolute and signed errors in separate LME models with PHASE (Pre or Post Training), TRAINING (Spatial or Non-spatial) and AZIMUTH as fixed effects. As before, we included PARTICIPANT (intercept and slope) and testing SESSION (intercept) as random effects.

The results of these analyses are also shown in Figure 27 (bottom row, panels B and C). We found a main effect of PHASE, caused by performance improvements after both training types (absolute error: $X^2(1) = 76.45$, $p < 0.001$; signed error: $X^2(1) = 58.07$, $p < 0.001$). Critically, the 2-way interaction between PHASE and TRAINING also reached significance (absolute error: $X^2(1) = 19.18$, $p < 0.001$; signed error: $X^2(1) = 27.26$, $p < 0.001$; for the complete summary of results of the LME analyses see Supplementary Table S1). Before training, localization errors were comparable across training type (absolute error: $t = 0.38$, $p = 0.70$; signed error: $t = 0.23$, $p = 0.82$). After training, errors decreased more substantially after the Spatial compared to the Non-Spatial training (absolute error: $t = 5.81$, $p < 0.001$; signed error: $t = 7.61$, $p < 0.001$). For signed error, also the 3-way interaction reached significance ($X^2(1) = 7.18$, $p = 0.007$): the impact of azimuth sound position on localization responses (i.e., larger leftward biases for sounds more toward the plugged side) decreased after the Spatial training ($X^2(1) = 23.26$, $p < 0.001$) but not after the Non-spatial training ($X^2(1) = 0.24$, $p = 0.62$). Mean with standard deviations for each of the conditions is shown in Table 5.

Localization errors in elevation remained unchanged, irrespective of training type. We noted only an upward bias in the post training session compared to the pre-training one (see mean \pm SD in Table 5 and 6).

		Spatial		Non-Spatial	
		Pre	Post	Pre	Post
Azimuth	Absolute error	19.3° \pm 14.1°	11.5° \pm 9.6°	19.1° \pm 14.4°	14.5° \pm 12.1°
	Signed error	-15.1° \pm 5.7°	-5.2° \pm 2.5°	-15.3° \pm 6.4°	-10.2° \pm 6.5°
Elevation	Absolute error	9.2° \pm 4.2°	9.5° \pm 5.4°	8.6° \pm 4.4°	9.2° \pm 5.1°
	Signed error	5.5° \pm 6.2°	6.5° \pm 7.8°	3.9° \pm 7.0°	6.0° \pm 6.9°

Table 5 STUDY 4 - PERFORMANCE IN HEAD-POINTING SOUND LOCALIZATION
Mean \pm SD for absolute and signed errors in azimuth and elevation as a function of PHASE (pre or post training) and training (Spatial or Non-Spatial during monaural listening). Results are pooled irrespective of the order in which the two training tasks were executed.

	X ²	df	p
AZIMUTH Absolute error			
testing phase	76.45	1	<.001
training type	0.14	1	.071
azimuth sound position	20.09	1	<.001
testing phase* training type	19.18	1	<.001
testing phase* azimuth sound position	7.78	1	.005
training type * azimuth sound position	5.39	1	.02
testing phase* training type * azimuth sound position	3.39	1	.07
AZIMUTH Signed error			
testing phase	58.07	1	<.001
training type	0.05	1	.082
azimuth sound position	9.96	1	.002
testing phase* training type	27.26	1	<.001
testing phase* azimuth sound position	0.21	1	.65
training type * azimuth sound position	1.53	1	.22
testing phase* training type * azimuth sound position	7.18	1	.007
ELEVATION Absolute error			
testing phase	3.40	1	0.07
training type	3.78	1	0.05
testing phase* training type	0.46	1	0.50
ELEVATION Signed error			
testing phase	22.31	1	<.001
training type	12.67	1	<.001
testing phase* training type	2.68	1	0.10

Table 6 STUDY 4 - RESULTS OF THE LME ANALYSES OF AZIMUTH AND ELEVATION ABSOLUTE AND SIGNED ERRORS

5.4.4. Do improvements during the spatial training predict generalization effects?

We asked if improvements during the spatial training (Figure 26) predicted the observed generalization effects, as measured in head-pointing sound localization (Figure 27). To this aim, we correlated performance in the two tasks. As indicator of improvement in the Spatial training, we used the individual slope coefficient obtained from the LME model on absolute error (see Figure 27A). The higher the slope coefficient, the more the participant improved in performance during the Spatial training (for clarity, we expressed improvements as positive numbers by multiplying each slope by -1). As indicator of improvements in head-pointing sound localization, we calculated the error difference (z-normalized) before and after spatial training, separately for absolute and signed error (again, to express improvements in signed error as positive numbers we multiplied the decreasing bias by -1). The higher the error difference, the larger the training generalization effect.

A correlation between the two measures emerged. The more a participant improved during Spatial training, the greater the reduction in absolute ($R = -0.77$, $p < 0.001$) and signed error ($R = 0.57$, $p = 0.009$) during head-pointing sound localization (see Figure 28).

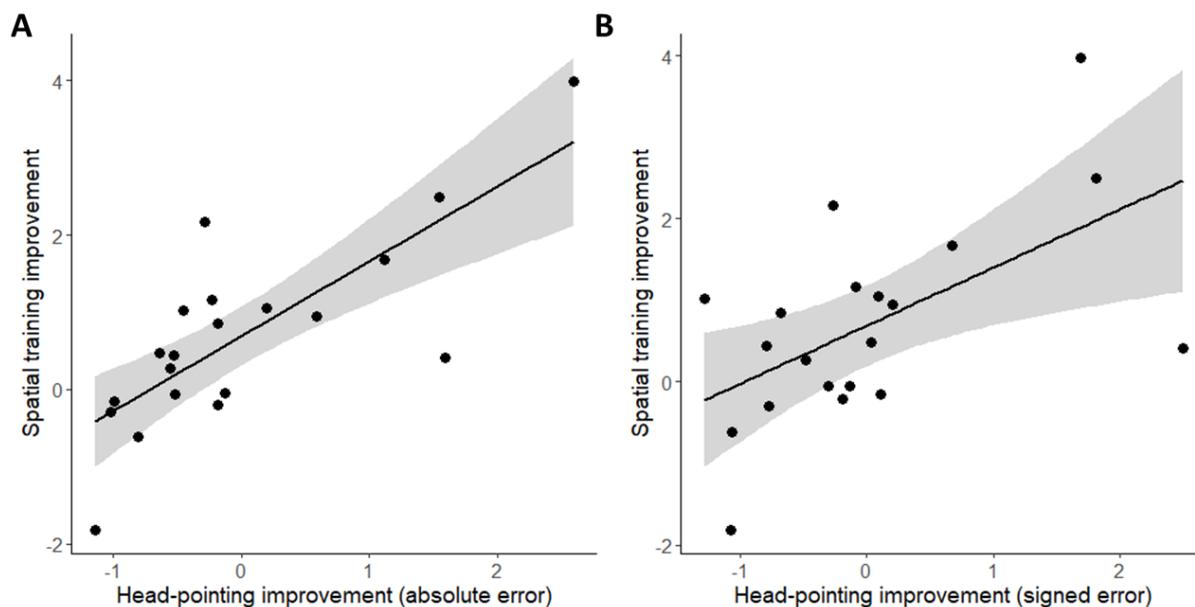


Figure 28 STUDY 4 - CORRELATION

Correlation between the improvement in spatial training and the improvement in head-pointing sound localization, computed on absolute (A) or signed (B) errors.

5.4.5. Does Spatial training change spontaneous head-rotation behavior while listening?

As anticipated in the Introduction, in our previous work (Valzolgher et al., 2020) we found that reaching to sounds increased head exploration movements during listening, and these head-movements were correlated with improvements in sound localization. In this section, we examined if our trainings also changed spontaneous head-movements during training (from sound onset until the first response but excluding the audio-visual feedback phase). To this aim, we examined three variables: (1) number of head-movements; (2) head-rotation extent; (3) head-rotation bias (positive values indicate rightward head-rotations). Recall that head-rotation extent and bias are referred only to movements around the vertical axis (see Methods).

Overall, participants made more head-movements in the Spatial (2.2 ± 0.8) compared to the Non-Spatial training (1.1 ± 0.4 , $t(19) = 5.94$, $p < .001$ on paired t-test). Furthermore, head-rotation extent was larger during the Spatial ($64.3^\circ \pm 25.9^\circ$) compared to the Non-Spatial training ($4.8^\circ \pm 1.8^\circ$, $t(19) = 10.17$, $p < 0.001$ on paired t-test). Finally, the head-rotation bias revealed that during the Spatial training participants rotated their head more toward the right ($21.3^\circ \pm 25.2^\circ$) as compared to the Non-Spatial training ($2.1^\circ \pm 4.5^\circ$, $t(19) = 3.21$, $p = 0.005$ on paired t-test). These findings indicate that spontaneous head-movements were elicited more in the Spatial compared to the Non-Spatial task.

To study if and how spontaneous head-movements behavior evolved during the Spatial training (i.e., reaching to sounds, as in Valzolgher et al., 2020), we entered the variables describing head-movements into separate

LME models using TRIAL NUMBER and SIDE as fixed effect, and PARTICIPANT (intercept and slope) and SESSION (intercept) as random effects. The main effects of TRIAL NUMBER and SIDE emerged for head-rotation bias (TRIAL NUMBER: $X^2(1) = 8.01, p = 0.005$; SIDE: $X^2(1) = 7671.55, p < 0.0001$) and head-rotation extent (TRIAL NUMBER: $X^2(1) = 3.82, p = 0.05$; SIDE: $X^2(1) = 292.83, p < 0.0001$), but not for number of head movements (TRIAL NUMBER: $X^2(1) = 0.003, p = 0.96$; SIDE: $X^2(1) = 1.27, p = 0.26$). As shown in Figure 29 A, participants increased their head rotation bias toward right (positive) as a function of trial repetition to approach the unplugged ear toward the sound direction. For head-rotation extent, also the 2-way interaction reached significance (TRIAL NUMBER X SIDE: $X^2(1) = 3.81, p < 0.05$, Figure 29 B). Finally, changes in head-rotation extent and head-rotation bias as a function of trials were correlated with one another ($R = 0.47, p = 0.04$).

In sum, during Spatial training phase, participants increased the explored space with the head and pointed more their heads toward the right as training progressed, particularly for sounds on the right (Figure 29).

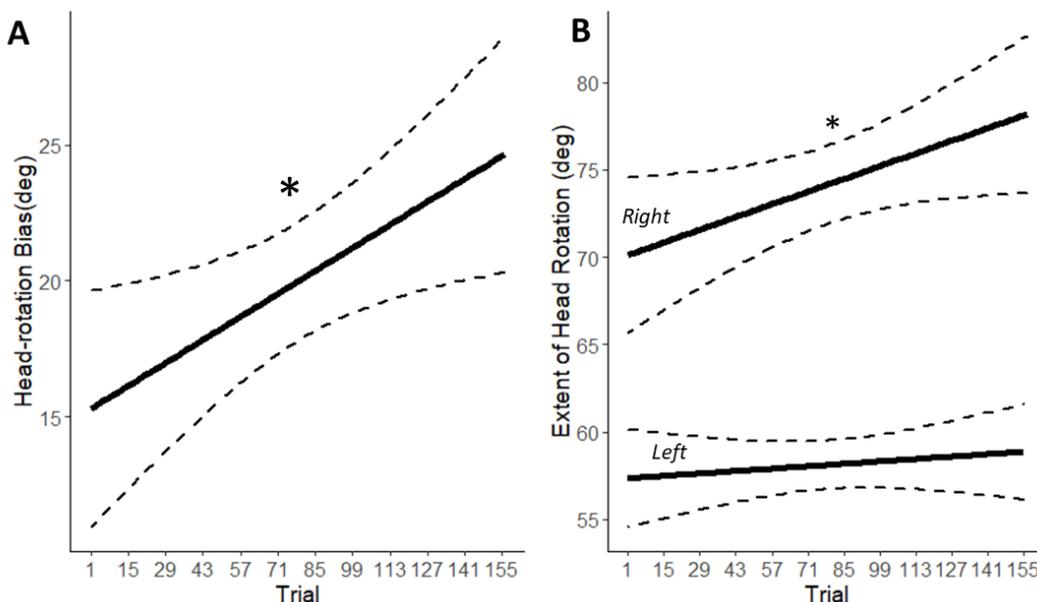


Figure 29 STUDY 4 - CHANGES IN HEAD BEHAVIOUR DURING THE SPATIAL TRAINING

Head-rotation bias as a function of trial (A) and extent of head-rotation as a function of trial and separately for sound delivered to the left or to the right of participant's midline (B). Dashed lines indicate the 95% confidence intervals of the overall linear regression.

5.4.6. Does training effects change immediate head-orienting to sounds?

In the head-pointing sound localization task, head-movements were functional to the response, hence they cannot be considered fully spontaneous (unlike in the training tasks). Yet, the first head-movement (which occurred around 0.99 seconds after the beginning of the sound) could be taken to reflect the immediate and spontaneous orienting response toward the sound. Here, we focused on these first movements to study if training effects can be detected also in immediate head-orienting to sounds. Specifically, we examined the horizontal direction of the first head-movement and its onset (i.e., Reaction Time, RT, in seconds).

An LME model on the horizontal direction of the first head-movement (with PHASE, TRAINING and SIDE as fixed effects and PARTICIPANT as random effect) revealed that first head movements were overall directed toward correct the hemispace (right: 46.2 ± 6.8 ; left: -42.0 ± 4.2 ; main effect of SIDE, $X^2(1) = 1270$, $p < 0.001$). Moreover, there were more rightward oriented responses after training (6.5 ± 2.9) compared to before training (-2.3 ± 1.7 ; PHASE, $X^2(1) = 8.82$, $p = 0.003$). This indicates that head-orienting to sounds was more biased towards the right (plugged) side after training, but irrespective of training type.

Most interestingly, a similar LME model on first head-movements RT (analyzed only for the trials in which the head-movement direction was correct, 97.1%) showed a 3-way interaction between PHASE, TRAINING and SIDE ($X^2(1) = 9.33$, $p = 0.003$). This interaction is illustrated in Figure 30. Before training, sounds delivered from the plugged right side resulted in longer first head-movements compared to sounds delivered from the left side, for both the Spatial (1.0 ± 0.2 vs. 0.9 ± 0.3 ; $p < 0.001$) and the Non-Spatial (1.1 ± 0.3 vs. 1.0 ± 0.3 ; $p = 0.02$) training. This difference in RT latency for the plugged side reversed selectively after the Spatial training (left: 1.0 ± 0.3 ; right: 0.9 ± 0.3 ; $p = 0.05$), whereas it persisted after the Non-Spatial training (left: 1.0 ± 0.3 ; right: 1.0 ± 0.3 ; $p = 0.009$). This finding provides further support to the notion participants were biased in orienting their heads toward the right (plugged) hemispace, but specifically after the Spatial Training.

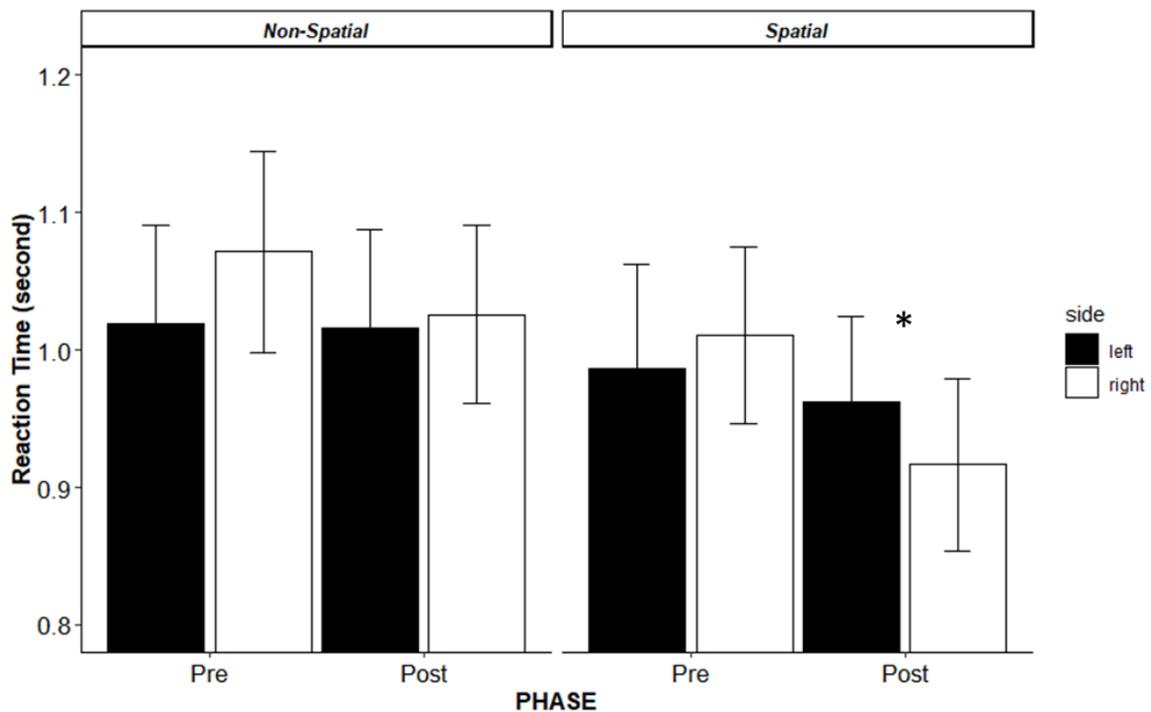


Figure 30 STUDY 4 - FIRST HEAD-MOVEMENTS REACTION TIME

Reaction Time, RT, in seconds as a function of training type (Non-Spatial or Spatial) phase (pre and post training), separately for sound delivered to the left (black) or to the right (white) of participant's midline. Error bars indicate the standard errors. In all comparisons RTs for right targets are significantly slower than those for left targets (see text), except in the post phase of the Spatial training. In the latter case the RT pattern is significantly reversed (marked by an asterisk).

5.4.7. Does training induce changes in audio-visual orienting?

Finally, we turned to investigate the effects of training in the audio-visual cueing task. To assess the immediate effects of monaural plugging on the audio-visual cueing task, we compared Cueing Effects (CE, in milliseconds) in binaural and monaural listening during the first session. We entered CE values in a LME model with the LISTENING (binaural or monaural) as fixed effect and PARTICIPANT (intercept) as random effects. As expected from previous works (Pavani et al., 2017), we found a main effect of LISTENING ($X^2(1) = 12.24, p < 0.001$) revealing that the CE decreased after monaural plugging.

To study the effects of our training protocols, we entered the CE values in monaural listening conditions into a LME model with PHASE (Pre or Post training), TRAINING (Spatial or Non-Spatial), and PARTICIPANT and SESSION (intercepts) as random effects. No main effects or interactions emerged (Figure 31, all $ps > 0.49$). This indicates that our spatial training affected performance during the training task itself, during head-pointing to sounds (both in terms of localization errors and first head-movements), but did not improve the implicit processing of sound side in relation to the visual stimulus.

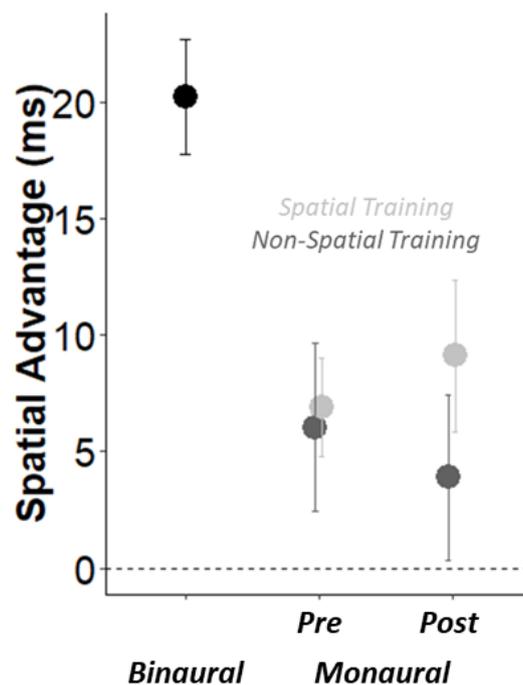


Figure 31 STUDY 4 - AUDIO-VISUAL CUEING TASK
Spatial advantage (ms) as a function of PHASE and TRAINING (Spatial in light grey and Non-Spatial in dark grey).

5.5. Discussion

Generalization of training effects is essential when assessing the potentials of any training procedure (Mendonça, 2014). Here we tested if a training based on sound-oriented actions can result in adaptations that

generalize to other sound localization tasks, that differ in terms of both stimuli and motor response. Using a crossover experimental design, we examined the effects of a Spatial training (indicate sound position through reaching) on two tasks performed before and after the training session: a head-pointing to sound task and an audio-visual cueing task. As a control, we used a Non-Spatial training (indicate sound type through reaching). Two main findings emerged. First, we confirmed that sound localization in listeners using an ear plug can rapidly improve during a Spatial training based on reaching to sounds (Valzolgher et al., 2020). Most importantly, we found that the improvement induced by the Spatial training, generalizes to head-pointing sound localization. Instead, the audio-visual cueing task remained unaffected. Second, we documented changes in head-movement behavior during the Spatial training. Most notably, we provide initial evidence of generalization of training effects can also impact on head-movement behavior in the head-pointing task, when focusing on the immediate head-movements to sounds. In the next paragraphs, we will discuss each of the main findings in turn.

5.5.1. Reaching to sound is an effective training strategy

In this study, we observed the generalization of training effects after the Spatial training. These results confirm and extend previous findings documenting the benefit of training based on audio-visual feedback (Rabini et al., 2019; Strelnikov et al., 2011), on head-pointing to sound (Steadman, Kim, Lestang, Goodman, & Picinali, 2019) and on motor interaction with sounds (Valzolgher, Campus, et al., 2020). In these previous works, however, the efficacy of training protocols was always examined using a between-subject experimental design. Participants' performance was measured before and after training, grouping participants as a function of the proposed training type. By contrast, in the present work, we proposed a within-participants experimental design. We successfully used a cross-over design to test generalization effects of a spatial training to sounds. This design gave us the opportunity to directly compare the effect of two different training situations in the same participants and to minimize any potential intergroup differences. Furthermore, such a cross-over design makes this paradigm particularly suited for clinical application because it permits to involve the totality of participants in the experimental training without the necessity of testing a control group (which cannot benefit from the training).

One additional feature of our study is the multifaceted nature of our training approach, which combines several of the intuitions that emerged from previous works. As proposed in the studies of Rabini and colleagues (2019) and Strelnikov and colleagues (2011), we exploited the benefit of providing audio-visual feedback about sounds' positions. In addition, we included goal-directed actions toward the sounds. A notable example in which motor action was used in combination with audio-visual stimulation to train acoustic space is the work of Steadman and colleagues (2019). They tested the ability to adapt to unfamiliar head-related transfer functions by involving listeners in training paradigms based on head pointing to sound and audio-visual feedback. The motor interaction with sound consisted in orienting the head to indicate sound direction. While this clearly constitute an active listening condition, it does not imply a direct interaction with the sounds. Hence, in our Spatial training, we built

on works in which the motor interactions with the sound sources occurred using the hand as effector, which have been proposed in some pioneering VR approaches (Honda et al., 2007; Ohuchi et al., 2005) as well as more recent studies (Poirier-Quinot & Katz, 2020).

This motor approach based on hand movements has been also adopted in our previous study (Valzolgher, Campus, et al., 2020), which demonstrated the effect of kinesthetic cues when re-learning to localize sounds with one ear plugged, as here. In this previous work, however, the action was largely passive and was not sound-oriented: participants were instructed to move their arm repeatedly while wearing a sound-emitting bracelet attached to their wrist. Conversely, in the Spatial training proposed in the current study, participants performed a reaching to sounds action.

Finally, a further distinctive element of our Spatial training was that participants were free to move their heads. Moving the head during training can improve localization because it introduces a dynamic dimension in the auditory cues (Brimijoin & Akeroyd, 2012; Pastore, Natale, Yost, & Dorman, 2018; Wallach, 1940). In addition, by moving the head people can implement adaptive strategies for listening: in our case, they moved the hearing ear to sample the occluded space with it. Steadman and colleagues (2019) also tested three training approaches based on head pointing to sound and audio-visual feedback. Interestingly, the training based on active listening, in which head movements were permitted during sound playback, resulted in greater improvements in localization accuracy as compared to the other two training versions they proposed (i.e., a non-gamified and gamified version of the training in which the head was still). In their study, however, there was only a limited time for exploring the auditory scene with head movements, because the sounds lasted only 1.6 seconds. Conversely, in our experimental training there was no time restriction on sound duration, and participants experienced the acoustic inputs if they preferred (approximately 4 seconds). This difference in sound duration may have permitted to take more advantage of the dynamics of the auditory cues and fostered more strategically behavior to identify the sound source.

Future studies should examine in more details the contribution of each of the components mentioned above in promoting adaptation to altered listening conditions. In particular, it would be relevant to understand which aspect(s) of the motor interactions with sound involved in the Spatial paradigm we proposed are most relevant. First, reaching to sound could enhance the spatial coding of sound source location because results from the coordination of different effectors (eyes, head, hand) into a common reference frame (Cohen & Andersen, 2002; Valzolgher, Verdelet, et al., 2020). Second, reaching could help directing attention toward the position occupied by the sound source, making the task more engaging (fundamental to foster relearning, see Dehaene, 2018). Third, reaching to sound could foster active listening through head exploration, as suggested by the documented shared reference frame between reaching motor command and head-orienting (Boyer et al., 2013).

5.5.2. A possible role for head-movements in adapting to altered spatial hearing

Our VR approach to spatial hearing (Verdelet et al., 2020; Valzolgher et al., 2020, b; Coudert et al., in press) allowed for the study of head-movements throughout the experiment. To this aim, we always presented sounds with long durations (i.e., until response, during the training phase; 3 seconds long, during the head-pointing task). Although the study of the contribution of head-movements to spatial hearing has been advocated since the 1940s, only few works have examined how head-movements could change while adapting to altered auditory cues (as here and in Valzolgher et al., 2020). In this respect our work adds to the limited body of work that focused on spontaneous head-movements during sound localization (Brimijoin & Akeroyd, 2012; Coudert et al., in press; Pastore, Natale, Yost, & Dorman, 2018; Wallach, 1940).

The analyses on head movements in our study resulted in two main findings. The first result concerns the Spatial training phase. Participants increased head-rotation extent, to explore a wider portion of space while listening to the sound. Furthermore, they turned their heads more to the right as the training trials progressed, as indicated by the increasing head rotation bias. These two head-orienting behaviors represents a spontaneous adaptation strategy to the altered binaural cues, which change spatial hearing specifically in the horizontal dimension. In particular, the rightward bias constitutes an effective strategy for directing the unplugged (left) ear toward the sound sources in both hemispaces. The second result concerns the post-training phase. In head-pointing to sounds, we observed again a bias in exploring the right (plugged) side of space. Participants directed their first head-movement (within the first second from sound onset) more to the right after training, and they triggered this rightward movement faster specifically after the Spatial training. This suggests that participants' head-related strategies implemented during the Spatial training transferred, to some extent, also to the first head-movement to sound post-training.

The adaptive behavior we documented during the Spatial Training suggest that participants adjusted their head-orienting strategies. This self-regulation behavior may be considered a form of cognitive offloading, that is a behavioral strategy that has the potentials to reduce cognitive load related to the increased perceptual complexity of the task (Risko & Gilbert, 2016). These strategies can emerge only if participants have the possibility of interacting with sounds, and are instead fully prevented when participants use a chin-rest or when sounds are delivered for a very short time. From this point of view, our results may be informative of the actual behavioral adaptations that listeners implement when coping with altered auditory cues. This approximation to daily life could open new directions of research to promote the processes of re-adaptation to altered auditory cues in clinical settings.

These findings have also the potentials to expand current interpretations of the mechanisms involved when learning to localize sounds with altered auditory cues. The focus until now has been on the neurocognitive

mechanisms of auditory cues reweighting and/or remapping. Cue-reweighting indicates the mechanisms by which the brain learns to weight the more reliable auditory cues when others are impoverished (Keating & King, 2013). For instance, when one ear is plugged the brain can learn to weight monaural cues more than the altered binaural cues, when localizing sounds in azimuth. Cue-remapping (Keating et al., 2015) refers to the ability to create new correspondences between auditory cues and spatial positions through experience. Our results show that self-regulation strategies related to head-behavior can also play a role when adapting to altered auditory cues.

One open question for future research – and a limitation of the present study – is the fact that we cannot establish to what extent the observed improvements should be exclusively attributed to changed head-movement behavior, or could also entail cue-reweighting and/or cue-remapping mechanisms. Two aspects of our findings may be in favor of the first account. First, our Spatial training did not impact the audio-visual cueing task, in which all head movements were prevented by means of chin-rest. Therefore, no advantage related to head-orienting strategies could emerge in this specific test. Future studies should address the contribution of head-movement to sound localization re-learning in a more direct manner, for instance by comparing two identical acoustical tasks to be performed before and after training. One task could permit head movements during listening (as in our head-orienting task) and the other should impose a head-still condition during listening (as in most auditory spatial studies in the literature). This design may help in disentangling if the generalization advantage we have observed is to be attributed exclusively to having learned a postural adjustment of the head.

A further hint supporting the idea that the generalization observed may be based on adjustments in head-movements comes from the study of the first movement of the head in the head-pointing to sounds task. This movement occurs on average in the first second after sound's appearance. The fact that the Spatial training speeded up rightward head-movements strongly suggests that the biased head-behavior developed during the training phase generalized to head-pointing to sounds.

5.6. Conclusions

Our findings show that reaching to sounds improves spatial hearing in participants using an ear-plug, more than a control condition using the same stimuli but non-spatial task demands. Training by reaching also modified head-movement behavior. Crucially, the improvements observed during training generalize also to a different type of sound localization task, possibly as a consequence of acquired and novel head-movement strategies. These findings widen the perspective about the mechanisms by which humans can adapt to altered auditory cues. In addition, they could have useful implications in clinical and applied settings. They demonstrate that implementing multisensory-motor approaches to train acoustic space perception is possible and future studies could extend this approach to hearing-impaired population with spatial hearing difficulties. For example, future

studies could test the effectiveness of reaching-to-sound training protocols by measuring the performance of deaf patients with unilateral hearing loss or people who wear cochlear implants or hearing aids (Litovsky et al., 2012).

Chapter 6. Study 5 - Reaching to sounds improves spatial hearing in bilateral cochlear implant users. A Randomized Clinical Trial.

6.1. Abstract

In this study, we assessed if spatial hearing training improves sound localization in BCI users and whether this effect can generalize to different sound localization tasks. Using a crossover randomized clinical trial, we tested in twenty BCI users the effect of a multisensory-motor spatial training on two different tasks: head-pointing to sound and audio-visual cueing. The spatial training involved reaching-to-sounds arm movements, to identify their position. Wrong responses were accompanied by audio-visual feedback. The sound lasted until correct response and spontaneous head-movements during listening were allowed and measured. As control, participant completed a Non-Spatial training, in which comparable reaching movements served to identify amplitude modulations in the sound. The stimuli and the multisensory nature of the feedback were always identical, irrespective of training type. In training and in the head-pointing to sound, we took advantage of a virtual reality and motor tracking approach permitting active listening. We found that BCI users reduced their sound localization errors in azimuth and adapted their spontaneous head-movements as a function of sound eccentricity during the spatial training. These effects generalized to a head-pointing sound localization task, determining a higher decrease of sound localization error in azimuth and more accurate first head-orienting response as compared to the control training. No training effect was observed on the audio-visual cueing task, but BCI users benefited from auditory spatial cues for orienting visual attention. These findings revealed that sound localization in BCI users improves with reaching-to-sound training, with benefits to a non-trained sound localization task. These findings have potentials for novel rehabilitation procedures in clinical context.

6.2. Introduction

Cochlear implants have become standard intervention for neurosensory deafness (Moore & Shannon, 2009; Wilson, 2019). Bilateral cochlear implants (BCI) are increasingly common, thus allowing partial recovery of interaural auditory cues (Aronoff et al., 2010; Gifford & Stecker, 2020). Benefits of CIs for speech understanding are now well-established in children (Allen, Nikolopoulos, & O'Donoghue, 1998; Calmels et al., 2004), adults (Stickney, Zeng, Litovsky, & Assmann, 2004; Wilson et al., 1991) and ageing individuals (Forli, Lazzarini, Fortunato, Bruschini, & Berrettini, 2019). In addition, binaural hearing improves sound localization in the horizontal dimension for BCI users, compared to unilateral ones (Smulders et al., 2016), or to conditions in which one of the two CI is switched off (Asp et al., 2012; Litovsky, Parkinson, & Arcaroli, 2009).

Yet, spatial hearing skills of BCI users are substantially reduced compared to those of normal hearing when locating stationary (Jones, Kan, & Litovsky, 2014; Kerber & Seeber, 2012) as well as moving sound sources (Moua, Kan, Jones, Misurelli, & Litovsky, 2019). A key factor causing such a poorer performance links to the reduced availability of auditory cues. Cochlear implants limit input resolution in the temporally and frequency domains, alter sound levels through gain controls, and change the auditory cues by effect of microphone filtering

strategies. In addition, the two CI processors work in isolation from one another, leading to independent acoustic streams to the brain. BCI users can thus access a limited set of auditory cues to localize sound, among which the most salient ones are the interaural level differences (ILDs) (Grantham, Ashmead, Ricketts, Labadie, & Haynes, 2007; Grieco-Calub & Litovsky, 2010; Seeber & Fastl, 2008; Van Hoesel, 2004).

All these limitations pertain to peripheral auditory signals. Can we improve spatial hearing abilities of BCI users by exploiting central neuro-cognitive strategies? (Isaiah & Hartley, 2015) Relearning spatial hearing skills is possible in deafened ferrets with BCI after a multisensory training with interleaved auditory and visual stimuli (Isaiah, Vongpaisal, King, & Hartley, 2014). Training-induced improvements were observed in the responsiveness of auditory cortical neurons and in their sensitivity to interaural level differences. Improvements emerged also in behavior: both in the final sound localization responses and in sound-evoked head-orienting movements made by early-deafened ferrets with BCI. Studies with normal hearing participants exposed to altered auditory cues also suggest that relearning spatial hearing is possible. Training protocols leveraging multisensory perception and active interactions with sounds proved effective in promoting adaptation to altered auditory cues and improved spatial hearing skills (Carlile, 2014; Keating, Dahmen, & King, 2015; Mendonça, 2014; Rabini, Altobelli, & Pavani, 2019; Strelnikov, Rosito, & Barone, 2011; Valzolgher, Verdelet, et al., 2020). Finally, preliminary results in UCI (Luntz et al., 2005) and BCI users (Tyler, Witt, Dunn, & Wang, 2010) also suggest that feedback-based training could improve sound localization also in this population. Yet, these pioneering works involved a limited number of participants and – critically – they did not explore if training effects could generalize beyond the trained task.

The aim of the present study was to assess whether spatial hearing can be trained in BCI users and whether training effects can generalize to different auditory tasks. Building upon our previous work in normal hearing adults listening with one ear plugged (Valzolgher, Verdelet, et al., 2020), we trained BCI users with a multisensory-motor task that combined reaching to near-by sounds, audio-visual feedback about sound position, and free head-movements. To assess training specificity, we used a crossover experimental design (Figure 1A; see Movie at https://youtu.be/D3ra1u_SBVM) in which we compared the experimental spatial training with a control training that used the same stimuli and multisensory feedback, but different task demands (i.e., a non-spatial discrimination of amplitude modulations in the sound). To assess training generalization effects, we measured participants' performance before and after each training type using two auditory tasks. The generalization tasks (a head-pointing sound localization task and an audio-visual cueing task, see below) differed with respect to the spatial training task both in terms of stimuli and response. Because our apparatus allowed for unrestrained but measurable head-movements, in line with Isaiah et al. (2014) we also studied changes in sound-related head-orienting movements. We predicted improvements in sound localization across trials during the spatial training (Valzolgher, Verdelet, et al., 2020). In addition, we expected greater

generalization effects after the spatial compared to the non-spatial training. Finally, we predicted changes in sound-related head-orienting movements during and after training (Valzolgher, Verdelet, et al., 2020) (Isaiah et al., 2014).

6.3. Materials and methods

6.3.1. Study Design

All participants (BCI users) took part in two experimental sessions, with at least 15 days of washout interval. Each session comprised two testing phases and one training (Figure 32 A); they were identical in all respects except for the training task. Participants performed both training types, which were counterbalanced across participants in a within-subject crossover design. In one session, they engaged in the experimental Spatial Training task; in the other, they performed the control Non-Spatial Training task. The testing phases included two auditory tasks: a head-pointing localization task (conducted in VR) and an audio-visual cueing task (conducted outside VR, see below).

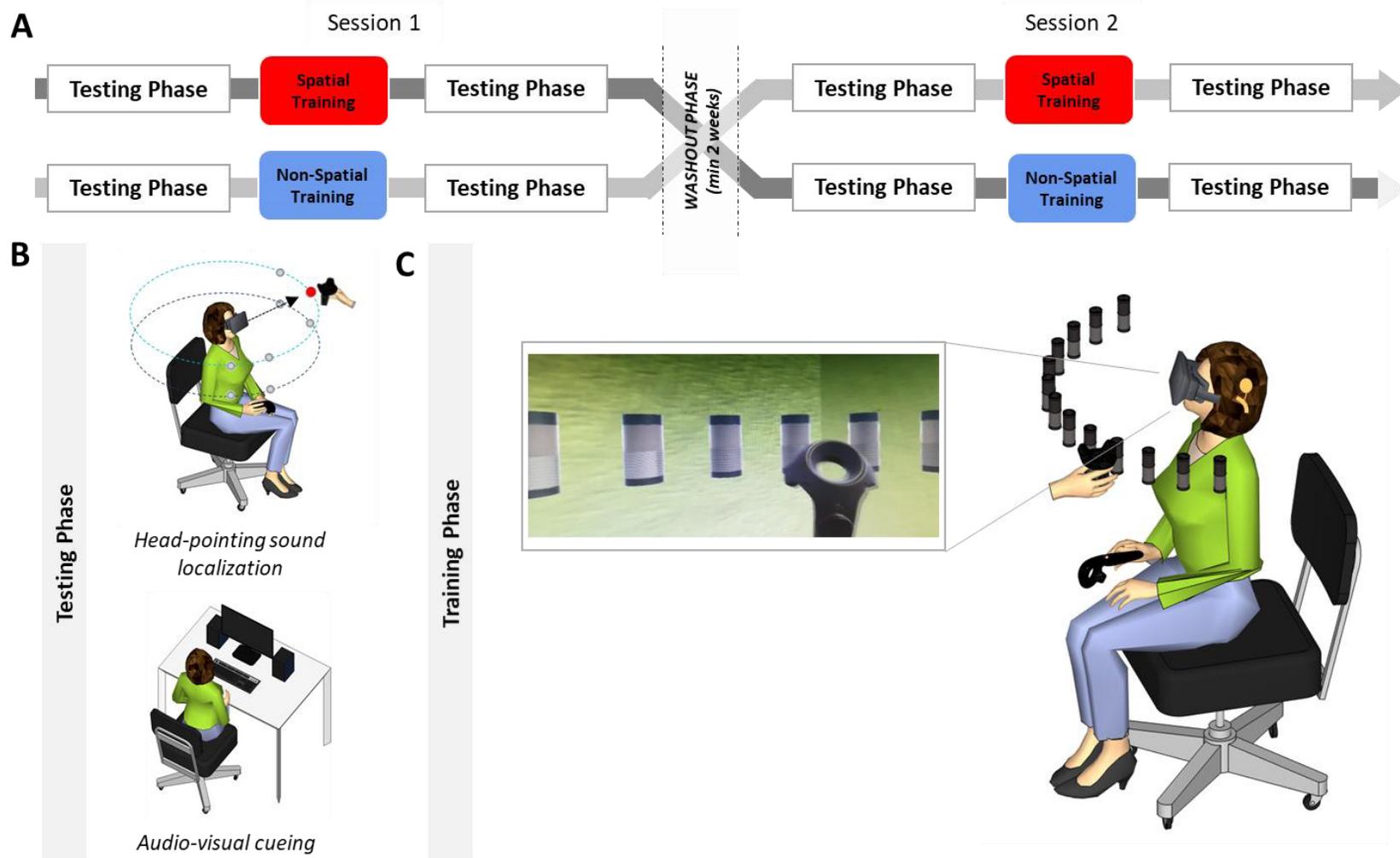


Figure 32 STUDY 5 - EXPERIMENTAL PROCEDURE AND SETTING

(A) Schematic description of the overall crossover design. Each session (Session 1 and Session 2) comprised by two testing phases, with a training task in between. (B) Testing Phases. Top: schematic representation of the participant wearing the HMD and holding the VR controller during the head-pointing sound localization task. The eight spheres in front of the participant indicate pre-determined speaker positions (not visible in the HMD). Bottom: schematic representation of the setting for the audio-visual cueing task (conducted entirely outside VR). (C) Training Phase. Schematic representation of the participant wearing the HMD and holding the VR controller. The cylindrical speakers were visible only in VR (see inset) and indicated the possible sound positions during both training types. Inset: close-up of the scene as visible inside the HMD from participant's perspective. During the task participant saw a virtual scenario comprising a room, the speakers and the VR controller they were holding in their hand. See Movie at https://youtu.be/D3ra1u_SBVM for animated description of the two training types.

6.3.2. Participants

Twenty BCI participants (age: M = 45.6, SD = 13.1, 7 males, 15 right-handed) took part in the study (see Supplementary Materials for power calculation). They were recruited at the ORL department of the civil hospital Edouard Herriot (HEH) in Lyon (France), and tested in a dedicated room inside the HEH premises. Participants signed an informed consent before starting the experiment, which was conducted according to the criteria of the Declaration of Helsinki (1964, amended in 2013) and it was approved by the Ethical Committee (Ile de France X, N° ID RCB 2019-A02293-54), and recorded in clinicaltrials.gov (NCT04183348). All had normal or corrected-to-normal vision and reported no movement or vestibular deficit, nor neurological or psychiatric history. Anamnestic and clinical data for individual BCI participants are provided in Table 1. During the experiment, participants used their daily sound processor settings (see Table S1 for details about CI model, processor strategy and microphone settings), to preserve the acoustic experience that each participant developed with the CIs.

ID	Gender	Age	Aetiology of deafness	Age at deaf diagnosis (y;m)	Age at first hearing aid (y;m)	Age at implantation (right ear)	Age at implantation (left ear)	Years with one CI	Years with two CIs
1	F	22	Unknown	0;5	0;7	2	15	13	7
2	F	65	Unknown	45	49	54	53	1	11
3	F	60	Genetic	10	10	40	51	11	9
4	F	46	Meningitis	1	5	39	44	5	2
5	F	47	Genetic	10	12	42	43	1	4
6	M	55	Genetic/ Sudden deafness	1	44	50	52	2	3
7	M	41	Ototoxicity	2	6	34	35	1	6
8	F	47	Genetic	16	21	42	39	3	5
9	M	32	Meningitis	2	2	3	28	25	4
10	M	25	Unknown	3	3	18	21	3	4
11	F	49	Unknown	13	30	39	39	0	10
12	M	59	Head trauma	52	No	53	53	0	6
13	F	47	Genetic	3	3	45	42	3	2
14	F	54	Genetic	42	42	50	51	1	3
15	M	24	Genetic	2;6	3	18	12	6	6
16	M	65	Otosclerosis	35	55	59	60	1	5
17	F	32	Otosclerosis of the cochlea	10	12	23	25	2	7
18	F	38	Genetic	13	18	26	30	4	8
19	F	51	Unknown	6	6	43	46	3	5
20	F	52	Genetic	6	23	43	51	8	1

Table 7 STUDY 5 - BIOGRAPHIC AND ANAMNESTIC AND CLINICAL CHARACTERISTICS OF CI PARTICIPANTS

6.3.3. Procedures

6.3.3.1. Testing Phases

Head-pointing sound localization task. This task was carried out entirely in VR, always using real sounds delivered in free field from pre-determined positions computed in each trial based on initial head-position (Coudert et al., 2020; Valzolgher, Alzhaler, et al., 2020; Valzolgher, Verdelet, et al., 2020). Participants were immersed in a virtual room that matched the size of the real one but was devoid of any object. Participants were informed that sound could be presented anywhere in the surrounding space. In each trial, a single sound (3 seconds long) served as target. During sound emission, head movements were unrestrained, thus participants were free to exploit in active listening strategies. They were instructed to point with their head toward the perceived position of sounds as soon as the sound finished and validate their response (Figure 32 B).

Audio-Visual Cueing task. This task was carried out entirely outside VR. Participants sat at a desk, in front of a computer monitor flanked by speakers. Their head-movements were restrained by a chin-rest. In each trial a single visual circle randomly appeared on the computer screen, in the left or right visual hemifield (10 degrees of eccentricity), above or below the visual meridian. Participants were instructed to indicate the elevation of visual disk (i.e., above or below eye level) as quickly and accurately as possible. Up/down responses were given using the up/down arrows keys on an external keyboard (Figure 32 B). Crucially, in each trial a task-irrelevant sound was presented simultaneously from one of the loudspeakers flanking the screen, either on the same side as the visual disk (spatially congruent trials), or from opposite side (spatially incongruent trials).

6.3.3.2. Training Tasks

Both training tasks were carried out entirely in VR. Participants were immersed in the same virtual room as the head-pointing sound localization task, but saw 13 virtual loudspeakers in their frontal space (see Figure 32 C and Movie). In each trial, a single sound was presented from one of the speakers. Sounds were delivered by bringing the real speaker into correspondence with one of the virtual ones, following a computer-controlled procedure based on continuous position tracking. Half of the sounds were amplitude modulated at 2 Hz, the remaining half at 3 Hz. Crucially, in both training tasks the stimulation was identical (i.e., sounds with two amplitude modulations were delivered from positions that changed on a trial-by-trial basis). The two training tasks differed only for the instructions given to participants.

Spatial Training. Participants were instructed to reach the speaker emitting the sound using the hand-held VR controller. The sound lasted until the participant reached and 'touched' the correct speaker. If they reached to the wrong speaker, the correct loudspeaker started to display concentric red beams that expanded from the correct position to reach (see Movie).

Non-Spatial Training. Participants were instructed to identify the amplitude modulation in the target sound, and indicate their discrimination through a reaching movement (again using VR controller). For fast amplitude-modulated sounds, participants were instructed to reach with the VR controller in front of them, aiming to touch an invisible virtual button above the central speaker. For slow amplitude-modulated sounds, participants reached instead an invisible virtual button below the same central speaker. As in the Spatial Training feedback procedure, the sound stopped only when a correct response was provided, and the same audio-visual feedback was displayed in case of wrong responses (see Movie).

For further details, see Supplementary Materials in the appendix at the end of the Chapter

6.3.4. Statistical Analysis

Linear mixed-effect modelling was used for all statistical analyses. Statistical analyses were run using R (version 1.0.143). For the linear mixed-effect (LME) model, we used the R-packages *emmean*, *lme4*, *lmerTest* in R Studio (Bates, Mächler, Bolker, & Walker, 2015; Fox & Weisberg, 2021). The R-package *car* was used to obtain deviance tables from the LME models.

6.4. Results

6.4.1. Changes in sound localization and head-movements during Spatial training

Sound localization error at the beginning of the study was 24.22 ± 15.82 (head-pointing sound localization before training in the first session), matching previous studies in documenting a low auditory spatial resolution, for example when compared to normal hearing performance (Jones et al., 2014; Kerber & Seeber, 2012).

To examine if sound localization improved during the Spatial Training we studied changes in absolute error in the horizontal dimension across successive trials using a LME model with trial number as fixed effect (Valzolgher, Verdelet, et al., 2020). To account for the variability related to individual participants and session in which the Spatial training was completed (first or second), we also included the *PARTICIPANT* (intercept and slope) and *SESSION* (intercept) as random effects in the model. As shown in Figure 33 A, BCI users reduced their absolute error across trials (main effect of trial number, $X^2(1) = 15.84$, $p < 0.001$). Performance in the Non-Spatial Training was near ceiling (mean error = 1.4%; see cumulative discrimination error across participant in Figure 33 B).

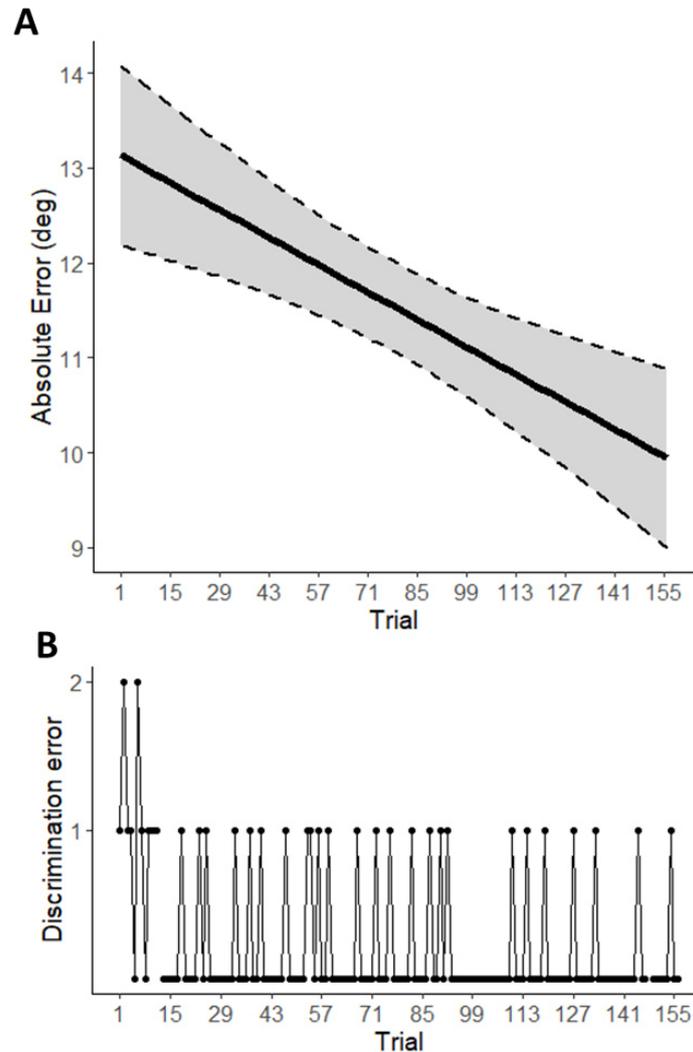


Figure 33 STUDY 5 - PARTICIPANTS' PERFORMANCE DURING TRAINING

(A) Reduction of absolute error in the horizontal dimension, as a function of trial in the Spatial Training. Linear regression (solid line), with 95% confidence intervals (dashed lines). (B) Cumulative discrimination error across participants as a function of trials in the Non-Spatial Training.

During training the head was unrestrained and the target lasted until correct response, allowing active listening behavior (i.e., spontaneous head movements). To study head movements, we extracted two dependent variables: number of head-rotations and head-rotation extent (Valzolgher, Verdelet, et al., 2020) (see supplementary materials for details). As expected by design, head-rotation patterns differed in the two training protocols: more head-rotations occurred in the Spatial (5.72 ± 3.24) compared to the Non-Spatial training (1.15 ± 0.41 , $t(19) = 5.72$, $p < .001$ on paired t-test); furthermore, the extent of head rotations was larger during the Spatial ($89.06^\circ \pm 41.86^\circ$) compared to the Non-Spatial training ($4.00^\circ \pm 1.91^\circ$, $t(19) = 9.20$, $p < 0.001$ on paired t-test). Interestingly, head-behavior evolved during the course of the Spatial training tasks. BCI users

progressively adapted their spontaneous head-movements as a function of sound eccentricity: the number of head-rotations diminished for central sounds (two-way interaction, $X^2(1) = 5.35$, $p = 0.02$), while head-rotation extent increased for more peripheral sounds ($X^2(1) = 4.78$, $p = 0.03$; see Supplementary Results and Figure 35).

6.4.2. Effects of Spatial training on head-pointing sound localization

Figure 34 A shows absolute localization errors in head-pointing sound localization across the two sessions of the experiment. The black line and circles indicate BCI users who performed the Spatial training before the Non-Spatial training; the grey line and circles indicate BCI users who completed the training in reversed order. The Spatial training improved performance to a greater extent compared to the Non-Spatial training, irrespective of the session in which it was completed.

To study the effect of Spatial training on head-pointing sound localization, we entered the absolute error in azimuth into a LME analysis with PHASE (Pre or Post Training), TRAINING (Spatial or Non-spatial) and AZIMUTH as fixed effects. We also included PARTICIPANT (intercept and slope) and testing SESSION (intercept) as random effects, to account for the variability across participants and the order of training type. We found a main effect of TRAINING ($X^2(1) = 4.89$, $p = 0.03$). Crucially, the 2-way interaction between PHASE and TRAINING reached significance ($X^2(1) = 24.14$, $p < 0.001$). The absolute error in azimuth decreased after the Spatial (Pre: 25.19 ± 6.18 ; Post: 19.31 ± 2.09 ; $t = 5.47$, $p < 0.001$; Figure 34 B-C), but not after the Non-spatial training (Pre: 22.84 ± 3.84 ; Post: 24.50 ± 1.29 , $t = 1.48$, $p = 0.14$; Figure 34 D-E).

Following Isaiah et al. (2014) we studied saccade-like head-orienting movements (i.e., movements occurring around 0.90 seconds after the beginning of the sound), which reflect the first spontaneous orienting response toward the sound. Saccade-like head-orienting response were more accurate after the Spatial Training, as revealed by a significant interaction between TRAINING and PHASE on response density at target location ($F(1,19) = 6.52$, $p = 0.02$, $n^2 = 0.11$; see Supplementary Results and Figure 36).

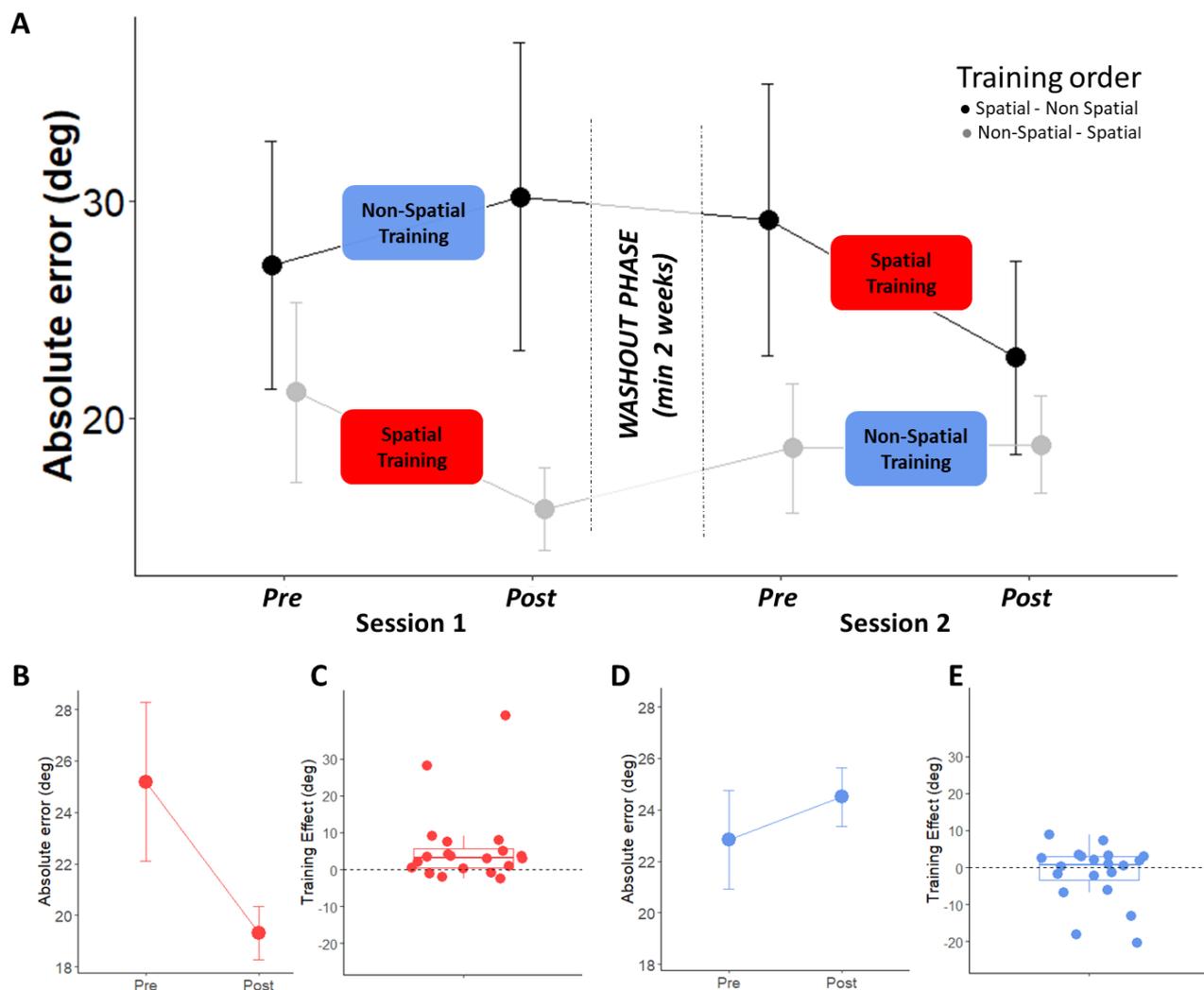


Figure 34 STUDY 5 - PERFORMANCE IN HEAD-POINTING SOUND LOCALIZATION

A) Absolute localization across the four testing sessions of the experimental design, separately for participants who completed the Spatial training on session 1 (grey) or session 2 (black line). B-D) Absolute errors (in degrees) as a function of PHASE before and after the Spatial (red, B) and Non-Spatial (blue, D) training. C-E) Training effect (deg): the difference between pre and post training session (Pre-Post) in absolute error as function of training task (C, Spatial in red and E, Non-Spatial in blue). Positive values in C and E represent improvements after the training.

To further investigate if these training benefits were modulated by clinical variables, we correlate the training effect (difference between in absolute error before and after the Spatial training) with the age at deafness diagnosis, years with one CI and years with two CIs. No significant correlation emerged (all p s > 0.35).

For completeness, we also analyzed absolute errors in elevation using a LME analysis with PHASE (Pre or Post Training), TRAINING (Spatial or Non-spatial) and ELEVATION as fixed effects. Again the 2-way interaction between PHASE and TRAINING reached significance ($X^2(1) = 16.75, p < 0.001$), caused by smaller vertical errors after the Spatial compared to the Non-Spatial training. The study of actual responses (instead of response error) showed that head-pointing in elevation was more biased towards zero (i.e., ear level) after Spatial compared to Non-spatial training (Spatial: 5.09 ± 6.73 ; Non-Spatial: 9.23 ± 1.43 ; $t = 6.05, p < 0.001$), whereas no difference emerged before training (Spatial: 7.55 ± 9.07 ; Nonspatial: 6.84 ± 6.76 ; $t = 1.03, p = 0.30$). This suggests that

the inability of BCI participants to discriminate elevation associated with the repeated exposure to sounds delivered at ear-level during the Spatial Training produced a bias in pointing (note that zero elevation corresponds to ear-level).

6.4.3. Audio-Visual Cueing task

To examine the effects of our training protocols on the audio-visual cueing task, we entered the response time (milliseconds) into a LME analysis with TESTING PHASE (pre or post training), TRAINING TYPE (spatial or non-spatial) and CONGRUENCY between sound and visual target position (Congruent or Incongruent). Again, PARTICIPANT NUMBER and SESSION (intercepts) were treated as random effects.

We found a main effect of CONGRUENCY ($X^2(1) = 30.45, p < 0.001$) caused by faster visual discrimination responses when the auditory cue matched target side (congruent: $M = 291.2$ ms, $SD = 78.9$ ms) compared to when it occurred on the opposite side of space (incongruent trials: $M = 307.3$ ms, $SD = 79.6$ ms). We also found a 3-way interaction between TESTING PHASE, TRAINING TYPE and SOUND POSITION ($X^2(1) = 4.90, p = 0.03$). After the Non-Spatial training, audio-visual cueing effect decreased (Pre: $M = 23.1$ ms, $SD = 17.8$ ms; Post: $M = 10.9$ ms, $SD = 20.5$ ms; $X^2(1) = 4.25, p = 0.04$), while it persisted after the Spatial one (Pre: $M = 11.9$ ms, $SD = 17.5$ ms; Post: $M = 18.4$ ms, $SD = 19.6$ ms; $X^2(1) = 1.15, p = 0.28$).

6.5. Discussion

A long-standing question in cochlear implant research is whether training can extend current guidelines for CI candidacy, by promoting (re)learning of auditory skills when using these devices (Fu & Galvin, 2007; Isaiah & Hartley, 2015; Isaiah et al., 2014; Rayes, Al-Malky, & Vickers, 2019). In the present study, we set out to investigate this question in human adults who received BCI, by studying the effects of a multisensory-motor training on sound localization abilities. Our findings demonstrated that BCI users can improve their horizontal sound localization across trials, while training in a reaching-to-sound paradigm with audio-visual feedback and their head free to move. Most importantly, the effects of this training generalize to horizontal sound localization in a different auditory localization task, using different sound positions and different response (head-pointing instead of manual reaching), but identical auditory stimuli.

Our findings extend previously documented training effects in animal models (Isaiah et al., 2014) and in normal hearing human listeners (Rabini et al., 2019; Steadman, Kim, Lestang, Goodman, & Picinali, 2019; Strelnikov et al., 2011; Valzolgher, Campus, Rabini, Gori, & Pavani, 2020). Most notably, they go beyond early reports on training effects in CI users (Luntz et al., 2005), by using a well-controlled crossover design which allows within-participant comparisons, and by showing generalization of training effects. It remains to be ascertained how long this improvement may persist over time. A preliminary indication from participants who performed the Spatial

training in the first session (N=10) shows that the error reduction induced by training persisted in this group after the 2-weeks washout ($X^2(1) = 4.64$ $p = 0.03$). Finally, to probe generalization effects when implicit sound localization is required, we also tested participants in an audio-visual cueing task. This task was not affected by the Spatial training. It is worth noticing, however, that lateralized auditory cues were found to be efficient in orienting BCIs' visual attention. This finding is in sharp contrast to previous work in UCI users, where this fundamental, cross-modal attentional ability was absent (Pavani et al., 2017). Interestingly, the audio-visual cueing advantage observed in BCI persisted after the Spatial, but not after the Non-Spatial training. This finding was unexpected, but it may reflect a bias for processing non-spatial rather than spatial sound attributes following the control training procedure.

A further contribution of the present study concerns head-movements behavior. Recent evidence suggests that head-movement related strategies may play a role in relearning sound localization abilities (Valzolgher, Verdelet, et al., 2020). Compared to static listening conditions, moving the head while listening to sounds improve sound localization in normal hearing (Brimijoin & Akeroyd, 2012; Hadley, Brimijoin, & Whitmer, 2019; Honda et al., 2013; McAnally & Martin, 2014; Perrett & Noble, 1997b, 1997a; Pollack & Rose, 1967; Vliegen, Van Grootel, & Van Opstal, 2004; Wightman & Kistler, 1999) and in BCI users (Coudert et al., 2020; Mueller, Meisenbacher, Lai, & Dillier, 2014; Pastore, Natale, Yost, & Dorman, 2018). Here, we show that head-behavior in BCI users changed across trial repetition during the Spatial training. Although this was unrelated to changes in performance, this preliminary observation suggests that BCI participants may also adapt their head-rotation behavior when re-learning their spatial hearing skills, by making it more purposeful across trials. Future studies on spatial hearing re-learning could focus on spontaneous self-regulatory behaviors during active listening, by adopting unconstrained head-movements and long-lasting sounds, as here.

6.6. Conclusion

Using a novel training approach based on reaching-to-sounds, audio-visual feedback and free head-movements during listening we show that training auditory skills in BCI users is possible. This improvement can be accounted for as a neurocognitive process (Wilson, 2017), since the auditory input from the CI remained unchanged. Furthermore, it stresses the positive contribution that multisensory and motor processes can play in adapting to CI (Anderson, Wiggins, Kitterick, & Hartley, 2017). Crucially, these findings have important implications for translational medicine. Broadening the perspective of training approaches from speech perception alone (Fu & Galvin, 2007; Rayes et al., 2019) toward spatial hearing, has the potential to extend current guidelines for CI candidacy and enrich the domain of actions that clinicians and patients themselves can undertake to improve their experience with the CIs.

6.7. Supplementary materials

6.7.1. Supplementary materials and methods

6.7.1.1. Participants

The choice of recruiting 20 participants was motivated by our recent study in normal hearing participants with one ear plugged, in which we run a protocol identical to the one applied here to BCI users precisely to gain further insights into the efficacy of our approach (Study 4). The effects size measured for generalization effects in this normal hearing population was 0.50 (reduction in absolute localization error before and after the reach-to-sound training). With a total sample size of 20 participants this results in an estimated power of 0.90. Further details about the cochlear implants used by BCI users are available in Table 8.

ID	Right ear						Left era					
	Year of implantation	Brand	Processor	Strategy	Microphone directionality	Number of active/total electrodes	Year of implantation	Brand	Processor	Strategy	Microphone directionality	Number of Active/total electrodes
1	2	Cochlear	Cp1000	Ace	Omnidirectional	22/22	15	Cochlear	Cp1000	Ace	Standard/ Omnidirectional	22/22
2	54	Advanced bionics	Naida ci q90	Hires-sw/ fidelity 120	Omnidirectional	16/16	53	Advanced bionics	Naida ci q90	Hires-sw/ fidelity 120	Omnidirectional	16/16
3	40	Cochlear	Cp910	Ace	Omnidirectional	20/22	51	Cochlear	Cp910	Ace	Omnidirectional	21/22
4	39	Medel	Rondo 9782	Fs4-p	No (only one microphone)	12/12	44	Medel	Rondo 2 (7016)	Fs4-p	No (only one microphone)	11/12
5	42	Medel	Sonnet	Fs4	Omnidirectional	10/12	43	Medel	Sonnet	Fs4	Omnidirectional	11/12
6	50	Advanced bionics	Naïda c90	Hires optima.s	T_mic	14/16	52	Advanced bionics	Naïda c90	Hires optima.s	T_mic	16/16
7	34	Oticon medical	Saphyr sp	Crysalis xdp	Directional (front)	18/20	35	Oticon medical	Saphyr sp	Crysalis xdp	Directional (front)	15/20
8	42	Medel	Sonnet	Fs4-p	Omnidirectional	9/12	39	Medel	Sonnet 2	Fs4	Omnidirectional	12/12
9	3	Cochlear	Nucleus cp910	Speak	Omnidirectional	20/22	28	Cochlear	Nucleus cp910	Ace	Omnidirectional	22/22
10	18	Medel	Sonnet 2	Fs4	Omnidirectional	12/12	21	Medel	Sonnet 2	Fs4	Omnidirectional	12/12
11	39	Medel	Rondo	Fs4	Omnidirectional	12/12	39	Medel	Rondo	Fs4	Omnidirectional	12/12
12	53	Cochlear	Cp1000	Ace	Omnidirectional	22/22	53	Cochlear	Cp1000	Ace	Omnidirectional	22/22
13	45	Cochlear	Cp1000	Ace	Scan	22/22	42	Cochlear	Cp910	Ace	Scan	19/22
14	50	Cochlear	Cp1000	Ace	Omnidirectional	17/22	51	Cochlear	Cp910	Ace	Omnidirectional	22/22
15	18	Cochlear	Cp1000	Ace	Omnidirectional	21/22	12	Cochlear	Cp1000	Ace	Omnidirectional	22/22
16	59	Oticon medical	Saphyr	Crysalis xdp	Omnidirectional	15/20	60	Oticon medical	Saphyr	Crysalis xdp	Omnidirectional	17/22
17	23	Oticon medical	Saphyr sp	Crysalis xdp	Directional (front)	19/20	25	Oticon medical	Saphyr sp	Crysalis xdp	Directional (front)	19/20
18	26	Oticon medical	Saphyr sp	Crysalis xdp	Directional (front)	20/20	30	Oticon medical	Saphyr sp	Crysalis xdp	Directional (front)	11/20
19	43	Medel	Sonnet	Fs4	Omnidirectional	11/12	46	Medel	Sonnet	Fs4-p	Omnidirectional	12/12
20	43	Medel	Sonnet	Fs4	Omnidirectional	12/12	51	Medel	Sonnet 2	Fs4	Omnidirectional	12/12

Table 8 STUDY 5 - EXTENDED INFORMATION ABOUT PARTICIPANTS' COCHLEAR IMPLANTS

6.7.1.2. Apparatus and general setting

Virtual reality and motion tracking were implemented using a HTC Vive system (Vive Enterprise). The system comprised one head-mounted display (HMD, resolution: 1080 x 1200 px, Field Of View (FOV): 110°, Refresh rate: 90 Hz), 2 hand-held controllers (one held by the experimenter to calibrate head-position, and one held by participants to interact with the virtual environment during the testing phase), 1 tracker mounted above the speaker to track its position in real time. Two lighthouse base-stations (Lighthouse V1.0) were also included for scanning the position of the controller and trackers. The position of controllers, tracker and HMD was monitored via a dead reckoning process using gyroscope and accelerometer, plus a correction signal from the lighthouse system every 8.333 milliseconds. Both tracking method allowed position tracking with a frequency sample of 250 Hz. Previous work from our group validated that tracking precision and accuracy of the HTC Vive System is adequate for behavioral research purposes (Verdelet et al., 2020). Stimuli were controlled and delivered using a LDLC ZALMAN PC (OS: Windows 10 (10.0.0) 64bit; Graphic card: NVIDIA GeForce GTX 1060 6GB; Processor: Intel Core i7-7700K, Quad-Core 4.2 GHz/4.5 GHz Turbo - Cache 8 Mo - TDP 95W) using Steam VR software and the development platform Unity3D (Unity Technologies, San Francisco, CA).

A distinctive aspect of our approach is that it brings real sounds into virtual reality scenarios, in a head-centered frame of reference. Free-field auditory stimuli were delivered using a loudspeaker (JBL GO Portable, 68.3 x 82.7 x 30.8 mm, Output Power 3.0W, Frequency response 180 Hz – 20 kHz, Signal-to-noise ratio > 80 dB), with the HTC Vive tracker firmly attached to its top surface. We continuously tracked the loudspeaker position, as well the position of the controller held in the participant's hand and the HMD. The software is designed to guide the experimenter to align the real loudspeaker (i.e., the sound source) with a set of pre-determined positions defined in the 3D virtual environment in each trial. This method combining virtual reality and motion tracking to measure sound localization abilities has been developed in our laboratory (Valzolgher, Alzhaler, et al., 2020; Valzolgher, Verdelet, et al., 2020). It was used for each of the training sessions and for head-pointing sound localization.

The Audio-Visual Cueing task was carried out without VR. The apparatus for this task included a dedicated PC and monitor (DELL 24"), a keyboard and two speakers. Speakers were positioned at ear level on both sides of the screen, at 20° from central fixation on either side (see bottom part of Figure 1B). The Audio-Visual Cueing task was implemented using OpenSesame®. Importantly, while performing this task participants' heads were supported using a chin-rest to prevent head movements.

All experiments were carried out inside a dedicated quiet room (3,6 m x 3,9 m x 2,7 m), with sound-proof panels applied to the walls to partially reduce echoes. The T60 reverberation of the room was 0,30-0,33 seconds, as measured using a Blue Solo 01bB phonometer. Before starting the experiment, participants

were instructed about the overall structure of the sessions, and accustomed to the VR equipment. When engaged in the VR tasks participants sat on a rotating armless chair with no chin rest, placed in the center of the room. They were also explained each task separately.

Data were collected during COVID 19 pandemic and thus experimenters and participants wore protective masks. To make sure participants understood the tasks well, instructions were given before they wore the helmet (and repeated if necessary during practice trials). Since it has been documented that surgical face-masks can affect communication (Giovanelli et al., 2020), if the participant requested it, the experimenter used a transparent face-mask to facilitate speech-reading during task explanation.

6.7.1.3. Testing Phase tasks

Head-pointing sound localization task. This task was carried out entirely in virtual reality (VR). The participant was immersed in a virtual room devoid of any objects, that reproduced exactly the size and shape of the real room. At the beginning of each trial, participants were asked to direct their gaze in front of them, by aligning their head with a central fixation cross. When the correct posture was reached, the fixation cross turned from white to blue. In the meanwhile, the experimenter placed the speaker in one of the possible eight pre-determined positions, using visual indications provided on a dedicated monitor. The pre-determined positions result from the combination of 4 different azimuths in front space (-67.5° , -22.5° , 22.5° or 67.5° ; with respect to the participant's body midline), 2 different elevations (10° above and below ear-level) and a single distance (55 cm from the center of the head; see top part of Figure 32 A). Crucially, the target sound was delivered only when the participant was in the correct start posture, and when the loudspeaker was positioned within a sphere with diameter of 5 cm centered on the pre-determined location. This ensured that each stimulation position was fully replicable at the beginning of each trial.

The target sound consisted of 3 second of white noise, amplitude-modulated at 2.5 Hz, and delivered at about 65 dB SPL as measured from the participant's head. Participants were informed that sounds could be delivered anywhere in the 3D space around them, and were instructed to indicate sound position using their heads as a pointer. Note that head and trunk movements remained unconstrained during and after sound emission, allowing spontaneous active listening behavior. Participants were free to move their heads and rotate the chair they sat on to explore the surrounding space. Importantly, they were informed that they were only allowed to validate their response after the sound ended, by pressing the button at the base of their hand-held controller. To emphasize this instruction, the fixation cross was white throughout the stimulation phase and turned red only when response validation was allowed. Since the speaker was invisible in VR, participants did not have visual cues about its position.

Each head-pointing sound localization task comprised 40 trials (5 repetitions, for each of the 8 sound positions) delivered in random order. Five practice trials were also introduced but discarded from the analyses. The task lasted approximately 15 minutes.

The head-pointing localization task differed with respect to training both in terms of spatial position of the targets and response effector (the head instead of the hand). The rationale was to minimize any potential effect of sensorimotor adaptation. The audio-visual cueing task was instead an implicit sound localization task, in which sounds serve as lateralized attention-orienting cues for the discrimination of visual targets (i.e., an audio-visual analogue to the classic attention cueing paradigm).

Audio-Visual Cueing task. This task was carried out entirely outside VR. Participants sat at a table placed inside the same room in which the VR experiments are carried out. Unlike the VR tasks, in this audio-visual cueing task participants rested their heads on the chinrest, hence head-movements were not allowed.

Each trial started with a white fixation cross appearing in the center of the monitor and remaining visible until response. After a random delay (450–600 ms), an auditory stimulus (white noise, duration 100 ms) was emitted from one of the two loudspeakers, at approximately 60 dB SPL as measured from head position. After 100 ms from sound delivery, the visual target was presented. This consisted of a white disk (20 pixels radius, 0.5° of visual angle), appearing on a black background for 140 ms in the upper or lower hemifield with respect to the horizontal meridian passing through visual fixation (1.15° above or below the meridian), either in the left or right hemifield (10° from fixation). In half of the trials, the visual target and the sound appeared on the same hemispace (congruent trials), whereas on the remaining half the visual target and the sound appeared on opposite hemispaces (incongruent trials).

Participants were asked to keep their gaze towards central fixation throughout the task and to indicate as quickly and accurately as possible the elevation of visual target. Up/down responses were given using the up/down arrows keys on an external keyboard, using the index/middle fingers of the right hand. Participants had a timeout of 2000 ms to give their answer and received feedback on accuracy (percentage of correct responses) and mean response time (in ms) only at the end of each block. Importantly, they were also explicitly informed that sounds were always entirely task-irrelevant.

The Audio-Visual Cueing task comprised 128 trials (16 repetitions, for each of the 4 disk positions and each sound side, divided in 2 blocks of 64 trials each) delivered in random order. Eight practice trials were also introduced but discarded from the analyses. The task lasted approximately 10 minutes.

6.7.1.4. Training tasks

The training phases were carried out entirely in VR and took place in the same virtual room used for the other tasks detailed above. During both training tasks, the participant saw 13 virtual loudspeakers inside the HMD. The loudspeakers were arranged in a semicircle centered on the head (radius 50 cm), at the ear level, and spanned between $\pm 72^\circ$ (with 12° of separation between each of them; see Figure 32 C).

The target stimulus consisted of a white noise, with two possible amplitude modulations (2 or 3 Hz). Targets sounds were delivered from each of the virtual speaker in random order (12 repetition for each of the 13 speakers, resulting in 156 trials overall, divided into 3 blocks of 52 trials each). Crucially, exactly the same stimuli were delivered in the two training tasks, thus making the auditory component of the trainings identical in this respect. In addition, both training tasks involved a similar motor response: a reaching movement. However, in the spatial training the reaching movements served to indicate the perceived sound position, whereas in the non-spatial training they served to indicate the perceived amplitude modulation in the target sound. Participants always held the controller with their right hand. At the beginning of each trial the controller was positioned on their sternum.

Spatial Training. Participants were instructed to reach the sound emitting speaker with the controller. The controller produced a brief vibration when coming into 'contact' with one of the virtual speakers, whether the response was correct or incorrect; hence it did not serve as feedback. Nonetheless, it served to enhance the feeling of interacting with a physical object (Valzolgher, Verdelet, et al., 2020). To prevent participants from colliding with the real speaker (held by the experimenter), the virtual speakers were presented closer to them, with a 5 cm offset in depth.

When participants reached and touched the correct speaker, the sound stopped. On the contrary, if they reached the wrong speaker, the correct speaker started flashing while the sound continued to play. Specifically, a series of red concentric circles (1024x1024 px, 2 circles per second) expanded intermittently, irradiating from the correct sound source. The rationale for using wide concentric circles was to capture the participants' attention even in case they were looking away from the correct sound source (e.g., toward the opposite hemisphere).

The visual feedback and the target sound stopped only when the subject reached the correct speaker position with the controller. This has two implications: first, a sense of agency was associated with the correct response (Valzolgher, Verdelet, et al., 2020); second, whenever the wrong speaker was originally selected, a combination of visual and auditory signals guided the participant to the correct sound source.

Non-Spatial Training. Participants were instructed to identify whether the target sound contained a fast (3 Hz) or slow (2 Hz) amplitude modulation. For fast amplitude-modulated sounds, participants directed the

controller in front of them, aiming to touch an invisible virtual button above the central speaker. For slow amplitude-modulated sounds, participants reached instead an invisible virtual button below the same central speaker. The position and effects of these virtual buttons was demonstrated to participants during the practice trials. A vibration from the controller always indicated that one of two buttons was correctly reached.

As in the Spatial Training feedback procedure, a visual feedback was delivered in case of wrong responses. This was a series of red concentric semicircles that expanded intermittently from above or below the central speaker to indicate the correct response. Recall that target sounds were presented from different spatial positions also during the Non-Spatial Training, but this spatial information was totally task-irrelevant.

6.7.2. Supplementary Results

6.7.2.1. Head-movements

In our task the head was free to move, both during training and during head-pointing to sound. Previous work from our group, using a similar active listening approach, has documented changes in spontaneous head-rotation behavior while re-learning spatial hearing abilities (Valzolgher, Verdelet, et al., 2020) and as a function of listening context (Valzolgher, Alzhaler, et al., 2020). Thus, we examined to what extent spontaneous head-rotations around the vertical axis changed during and after training in BCIs.

For the study of head-movements, we calculated the tangential velocity of the head using the combined displacement across all axes (expressed in degrees of rotation), with a 2-points central difference derivative algorithm and a 10-points half-window (for smoothing purposes). Onset and end of each movement were then computed automatically, using a velocity-based threshold ($10^\circ/\text{s}$) (Valzolgher, Verdelet, et al., 2020). Furthermore, each head-movement was checked manually by visualizing changes in rotation and speed using a custom-made toolbox in MATLAB R2018b. The rationale was to eliminate trials with artefacts or missing data points.

6.7.2.2. Training task

We examined head-rotations around the vertical axis and extracted two dependent variables: number of head-rotations and head-rotation extent. To calculate number of head-rotations, we counted the number of velocity peaks detected in each trial until the first response. Head-rotations smaller than 2° degrees were removed from the count to avoid the inclusion of micro-postural movements unrelated to the task. To calculate head-rotation extent, we summed the maxima of head-rotation on either side of space (left or right). For instance, if the head rotated 20° to the right and 40° to the left, the head-rotation extent was calculated as the sum of the two, hence 60° .

As anticipated in the main text, BCI users adapted their spontaneous head-movements as a function of sound eccentricity in two ways. To investigate if spontaneous head-movements changed during the Spatial training, we entered number of head-movements and head-rotation extent in separate LME models with TRIAL and SOUND ECCENTRICITY (central or peripheral targets) as fixed effect and PARTICIPANT (intercept and slope) and SESSION (intercept) as random effects. Across trial repetition, BCI users adapted their spontaneous head-movements as a function of sound eccentricity in two ways: number of head-rotations reduced for central sounds (two-way interaction, $X^2(1) = 5.35$, $p = 0.02$; Figure 35 A), while head-rotation extent increased for more peripheral sounds ($X^2(1) = 4.78$, $p = 0.03$; Figure 35 B). Changes in number and extent of head rotation correlated with one another, irrespective of sound eccentricity ($R = 0.82$, $p < 0.001$; others $p > 0.10$; Figure 35 C). The more the number of head-rotations increased during the Spatial Training, the more the space explored with head rotation also increased.

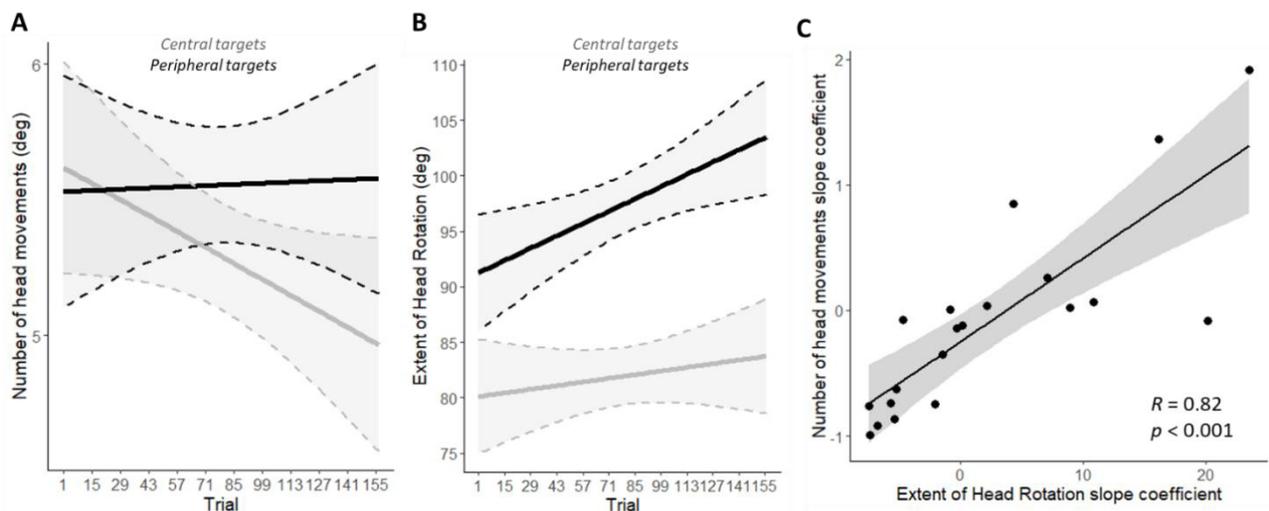


Figure 35 STUDY 5 - HEAD MOVEMENT IN THE SPATIAL TRAINING

(A) Number of head-movements as a function of trial and separated for central (grey) and peripheral targets (black); (B) Extent of Head Rotation as a function of trial and separated for central (grey) and peripheral targets (black); (C) Inter-participant correlation between individual number of head-rotations and head-rotation extent change during training (as indexed by slope coefficients as a function of trial).

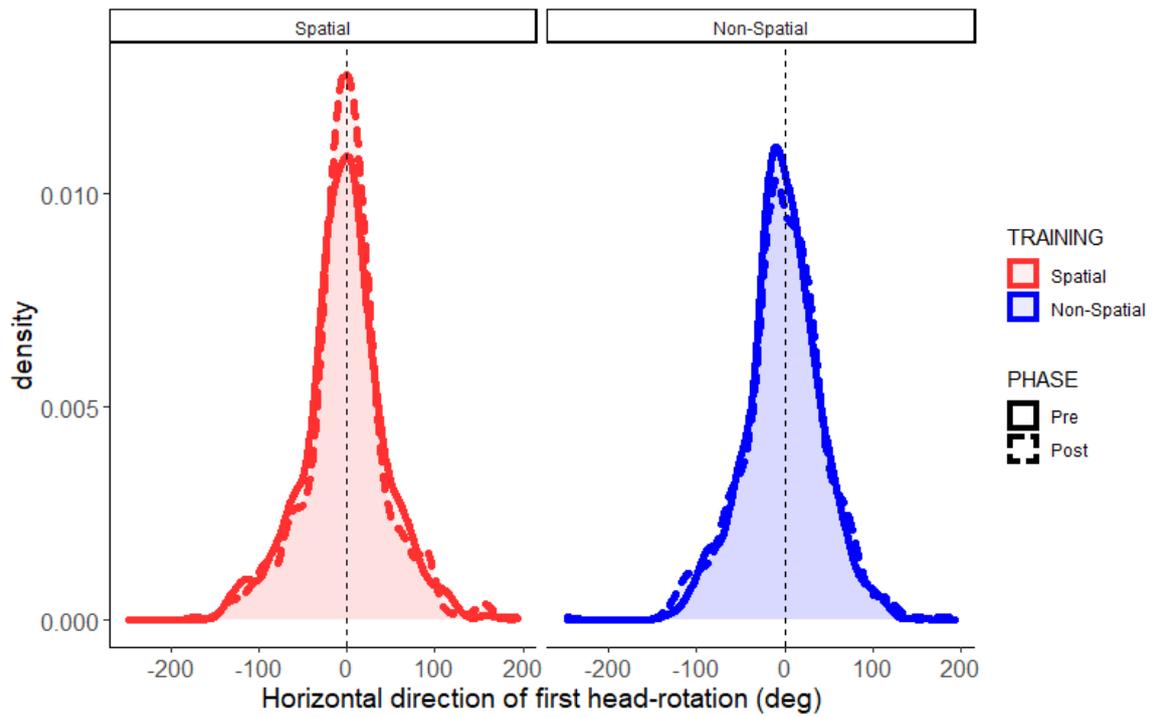


Figure 36 STUDY 5 - HEAD MOVEMENT IN IN HEAD-POINTING SOUND LOCALIZATION
 Distribution of the horizontal direction of the first head movement during the head-pointing to sounds task as a function of phase (pre and post) and training (Spatial and Non-Spatial).

Accordingly, we examined the distribution of horizontal direction of the first head-rotation. Figure plots response density before and after training and a function of training type. For clarity, we expressed head-pointing responses irrespective of target location (i.e., all positions in azimuth have been collapsed) and we performed an ANOVA on interparticipant density values at zero (i.e., actual target position). A significant interaction between training and phase emerged ($F(1,19) = 6.52, p = 0.02, \eta^2 = 0.11$), caused by more accurate first head-orienting response after the Spatial Training.

Chapter 7. General Discussion

7.1. Answers to the experimental questions

The overall aim of this thesis was to explore the cognitive mechanisms involved in the localization of sound and examine the role of active listening and sound-evoked actions in their localization. To these aims, I have used throughout the thesis an innovative virtual reality approach (VR). First, I measured the effect of visual and motor information on sound localization performance in young (Chapter 2) and ageing participants (Chapter 3). Then, I demonstrated the effectiveness of a reaching-to-sound training, based on multisensory feedback and active listening, in normal hearing people with one-ear-plugged (Chapter 4 and 5). Finally, I extended this training paradigm towards people with deafness that use bilateral cochlear implants (Chapter 6). In the Preview (page 20), I have listed the key experimental questions I aimed to address with this doctoral project in each of the five different experimental contributions proposed. In the following paragraphs, I briefly remind each question and the answers that emerged from the experimental studies.

7.1.1. Seeing an empty visual environment improves sound localization in elevation and produces faster gaze orienting responses to the sound

The study in Chapter 2 describes a first attempt to investigate the effects of seeing an empty and minimally structured environment on sound localization. In this study, I examined the impact of a minimal visual spatial frame on sound localization in normal hearing participants (N = 36). The hypothesis was that experiencing a structured visual environment could favor sound localization as compared to a uniform and unstructured grey scene. This hypothesis was guided by the notion that visual information could constitute a frame of reference that people exploit when extracting spatial information about sound position (Warren, 1940). The results confirmed the hypothesis and showed that sound localization improves in the presence of a minimal visual spatial frame, both in terms of accuracy in elevation and sound-induced gaze orienting behavior. This first experimental study served also the purpose to test an innovative method that exploits virtual reality and is particularly suited for investigating my research questions. The step forward compared to previous research concerns precisely the combination of a cutting-edge methodology with a research question that could only be addressed with this tool: presenting sounds all around participants but showing them only the visual background of the environment and not possible sound sources. A further interesting point is the use of the first movement of the head and eyes as dependent variables, as relevant descriptors of sound localization behavior. Studying sound localization behavior does not only mean extracting information about performance of participants but also investigating the implicit and immediate behavioral reactions evoked by a given sound or listening

situation. In this study, we documented that the participants oriented their gaze (eyes and head) towards the sound and did so with greater speed when immersed in a scenario with visual references.

7.1.2. Head movements improve sound localization, also in ageing people with presbycusis

The study in Chapter 3 examined whether spontaneous head-movements implemented during a sound localization task and a front-back disambiguation task improve localization performance of ageing adults with presbycusis. Hearing-impaired ageing adults (N=8) were tested in a front-back discrimination task and in a sound localization task in the frontal plane in two listening conditions. In the static condition, they were asked to keep their head still while listening to the sound; in the active condition, they were free to move the head while listening. We observed advantages in both tasks during active listening. When they implemented spontaneous head movements, performance increased significantly in front-back discrimination and localization errors decreased in the frontal plane, in azimuth and elevation, but not distance perception. This effect emerged despite head-movements were rather limited (3.2 movements per trial on average, 41.7° of head rotation around the vertical axis and 21.5° around the horizontal axis on average). This study agrees with the previous literature highlighting the potential of head-movements for spatial hearing. Further, it provided initial evidence of this advantage in ageing adults with presbycusis. These results emphasized the importance of having a naturalistic approach when testing this type of population: limiting active listening can lead to the underestimation of their real ability to interpret auditory cues for sound localization. Furthermore, this study proved that our VR approach can also be applied to elderly people with hearing loss – a population which is less used to VR equipment and scenarios. Finally, our work suggests that this documented benefit of active listening could be exploited when planning interventions in hearing-impaired populations.

7.1.3. Reaching to sounds helps adaptation to altered auditory cues more than naming the sources

The study in Chapter 4 tested the role of active multisensory-motor interactions in promoting adaptations to altered binaural cues in normal hearing adults with one ear plugged and muffled. By immersing participants in a VR scenario, this study showed that reaching to sounds improves monaural sound localization more than simply naming the label associated with the speakers. In both conditions, audio-visual feedback was provided and free-head movements during listening were allowed. In this study, two experimental groups were tested (N = 14 in each group) while performing a sound localization task that entailed either reaching the sources or naming them (between group comparison). Results show that both groups improved sound localization across trials. Interestingly, improvements were faster and larger in the reaching compared to the naming group. The benefits of reaching to sound were linked to

progressively wider head-movements for exploring the auditory space while listening. This head-movement behavior might be considered an example of strategic aspect of adaptation to altered auditory cues. In the adverse listening scenario (i.e. monaural listening), participants turned the head to achieve the most favorable position and, possibly, maximize the available auditory cues. This active strategy appeared to have been favored by the type of task: promoting an active motor interaction with sounds through reaching movements resulted in increased head-orienting behavior. This study documents for the first time that head-movements can play an important role in adaptation to monaural listening and it stresses the importance of fostering this strategy when implementing training protocols aiming at ameliorant spatial hearing.

7.1.4. The effects of sound localization by reaching generalize to a different sound localization task

In the study in Chapter 5 I tested if the effects of the previously proposed reaching-to-sound training can generalize to other untrained tasks that entail processing of sound position. I tested normal hearing participants plugged monaurally (N = 20) and I measured training effects in a head-pointing sound localization task and in an audio-visual cueing task before and after training. Using a crossover experimental design, I compared two training types within-participants: a Spatial and a Non-Spatial training. The Spatial training proposed a reaching-to-sound task identical to the one used in the Study 3; the Non-Spatial training, required participants to identify if the amplitude modulation in the sound occurred with higher or lower frequency and thus the sound spatial position was entirely task irrelevant. We delivered identical acoustic stimulation, but the different instructions made participants focus on either spatial or non-spatial features. This study demonstrated that reaching to sounds training induced improvements that generalize to head-pointing sound localization. By contrast, no training effect emerged for the audio-visual cueing task. Interestingly, I also replicated the observation on head-movement behavior change across trials when participants train by reaching to sounds. Finally, training change immediate head-orienting to sounds as revealed by the analysis on the first head-movement during the sound localization task suggesting that participants' head-related strategies implemented during the Spatial training transferred, to some extent, also to this task. Participants directed their first head-movement (within the first second from sound onset) more to the right after training, and they triggered this rightward movement faster specifically after the Spatial training. These finding extended what I observed in the Study 3 emphasizing first that training based on reaching to sound can generalize to a different task and second, they may expand current interpretations of the mechanisms involved when learning to localize sounds with altered auditory cues. The scholars' focus until now in this domain has been on the neurocognitive mechanisms of auditory cues reweighting and/or remapping; these results

show that active strategies related to head-behavior can and should also be considered part of this complex mechanism (see 7.2 for discussion).

7.1.5. Sound localization in bilateral cochlear implant users can be trained though reaching to sounds

In the study in Chapter 6, I applied the same training design developed in normal-hearing to a clinical population, for whom sound localization is often difficult: people with deafness using bilateral cochlear implants. This step towards the clinical context was the ultimate goal of this thesis. I recruited people with bilateral cochlear implants (N = 20) for a training task based on reaching to sounds, and tested both the direct and the generalization effects of this training using a paradigm identical to the one developed for the study in Chapter 5. I documented that training acoustic space of BCI users is possible using a reaching to sound based training paradigm. More interestingly, the generalization of training effect toward a head-pointing sound localization task emerged also in this population. Furthermore, testing BCI users in the audio-visual cueing task gave me the opportunity to observe that the information given by two cochlear implants is sufficient to restore audio-visual orienting of attention – unlike monaural listening with a single CI (see Pavani et al., 2017). Yet, this audio-visual orienting of attention was not further improved by the spatial training, although it was unexpectedly reduced by the non-spatial training. Interestingly, measuring head-rotation movements permitted to document changes across trials during training also in BCI users. This study represents the first systematic evidence that a motor and multisensory training paradigm can promote sounds localization also in BCI adults (see Isaiah et al., 2011 for related findings in ferrets with BCI). These findings may have implications for clinical practice and for personalized care and rehabilitation of the hearing-impaired (see the in-depth information box at page 174).

7.2. Further elements for the conceptual model of acoustic space relearning

“Apprendre, c’est donc saisir par la pensée : emporter en soi une parcelle de réalité, un modèle de la structure du monde” (“Learning is grasping a fragment of reality, catching it, and bringing it inside our brains”, Dehaene, 2018, page 23). In this sense, learning to locate sounds with new cues means extracting from the auditory reality a new spatial map that can be used for knowing where sounds are and interact with them. The processes determining the creation of a new spatial map involve both bottom-up multisensory processes and top-down multisensory mechanisms. Our brain can learn to localize sounds with new auditory cues (see section 1.2) and, so far, several crucial elements guiding this process have been identified. I believe the experimental works of this thesis adds to this model by showing a key role for visual input and spontaneous compensatory strategical behavior.

Relearning to localize sounds in real life situations involves the other sensory systems. Vision and touch contribute to create a complex sound source percept and they are crucial to extract spatial coordinates. Several studies in the literature (see Section 1.2.), as well as the ones proposed in the present thesis, took advantage of multisensory stimulation to train acoustic space perception. For instance, Strelnikov and colleagues (2011) proposed an audio-visual training for relearning spatial hearing and compared it to an auditory-only training. Performance improvements in the sound localization task were larger after the multisensory stimulation-based training compared to the auditory-only training. This finding was supported by previous and subsequent studies proposing also training based on multisensory feedback provided after participants' sound localization response (Shinn-Cunningham, Durlach, & Held, 1998; Zahorik, Bangayan, Sundareswaran, Wang, & Tam, 2006) and by multisensory training studies with ferrets (Isaiah, Vongpaisal, King, & Hartley, 2014). Providing complex stimuli involving different sensory modalities is crucial to favor relearning.

However, receiving multisensory inputs has often been treated as a passive exercise – a bottom-up flow of information in contexts in which listeners have little opportunities for exploring the environment. In this thesis, I have documented that active engagement in a task can play, in fact, an important role in favoring relearning. Being active when learning means having the opportunity to implement behavioral strategies such as head movement to make sense of reality. Consider, for instance, a classic visual experiment conducted in the 1960s by Held and Hein (Held & Hein, 1963). They tested the behavior of two kittens that lived for three hours per day, over a period of eight weeks, inside a cylinder entirely covered by drawings of vertical bars. Both cats received identical visual stimuli, but one was free to move actively and the other was passively moved around using a carousel mechanism (they could not locomote). Despite the poor visual stimulation provided (only vertical bars), the active cat presented higher visual exploration abilities (using three tests: visually-guided paw placement, avoidance of a visual cliff, blink to an approaching object) compared to the passive cat. Testing participants in localizing sound limiting their head movements may be considered comparable to what happened to the passive cat in the experiment conducted by Held and Hein. This setting limited the spontaneous exploration and consequently influenced the possibility to implement effective behavioral strategies to make better sense of the incoming sensory stimulation. In the studies proposed in this thesis, I have continuously adopted a pseudo-naturalistic method (see 7.3) aimed at giving listeners the opportunity to implement active behaviors. In particular, I focused on the possibility of interacting with nearby sound through reaching, and on head-rotations. Even these two minimal action contexts (note that participants were not free to move in the room to freely approach sounds) made sound localization into an active task – unlike most previous works in which listeners remained primarily static while experiencing the sounds and were allowed movements only during the response phase.

Conceptual models depicting the processes of subtending relearning of spatial hearing skills have been previously proposed. For instance, Mendonça (2014) in her review illustrated a model that captures the continuous updating of correspondences between the incoming auditory cues and external space coordinates. This auditory adaptation model describes the path of an input sound from its decomposition into auditory cues, to the extraction of a spatial coordinates in the perceived auditory space. It emphasizes the role of perceptual feedback accompanying the sounds: if the comparison between the percept and the feedback confirms the correspondence, the map is reinforced; if the feedback is different from the percept, a new cue combination rule is created (Figure 37).

Despite providing a useful schematization of the relearning process, this conceptual model does not describe the complexity of cues combination rules. To establish azimuth, elevation and depth perception, monaural and binaural auditory cues ought to be combined. For azimuth perception, binaural cues have a major contribution, while spectral cues are more relevant to extract elevation (see Section 1.1.1.). As described by Van Wanrooij and Van Opstal (2007), under perturbed hearing conditions, which cues are most reliable for the task may change (a process described as 'cues reweighting'). For instance, when participants are plugged monaurally, binaural cues become less certain for estimating azimuth and the weight of spectral cues at the unplugged ear increases. Hofman and Van Opstal (2003) suggested that such a weighting process occurs also for the spectral cues resulting from the two ears: monaural cues are processed independently and are subsequently associated through binaural weighting of the spectral cues. Therefore, the reweighting process is assumed to be a necessary step when creating new correspondence rules, but exists at different levels. In Figure 37, I have combined the original conceptual model proposed by Mendonça (2014), with the concept of cues reweighting expressed by other authors (Hofman and Van Opstal, 2003; Van Wanrooij and Van Opstal, 2007). To localize sounds' vertically, monaural cues of the ears are assembled and weighed together. In addition, they are further combined with the binaural cues, which are responsible for localization on the horizontal plane.

In addition to the process of reweighing auditory cues, it is also important to consider the concept of cues remapping. Remapping means being able to create through experience new correspondences between the auditory cues and the spatial positions in external space. As shown in the experiment of Hofman described in the Introduction (see 1.2.), humans are able to store the new correspondences created and the new maps stored do not interfere with each other (Hofman et al., 1998). In Figure 37 this concept is conveyed by representing different stored maps that can be used according to the available cues in which the reweighting processes change. However, it is important to note that this is a simplification. The remapping process could in fact take place separately for the binaural and monaural cues (Trapeau & Schönwiesner, 2015, Van Opstal, 2016). It is not yet clear how these two processes interact. In the model shown in Figure 37, I tried to represent that cues reweighting can take place without necessarily

determining the creation of a new map (same color, but different texture). While cues remapping implies a different process and the creation of different maps (drawn here with different colors).

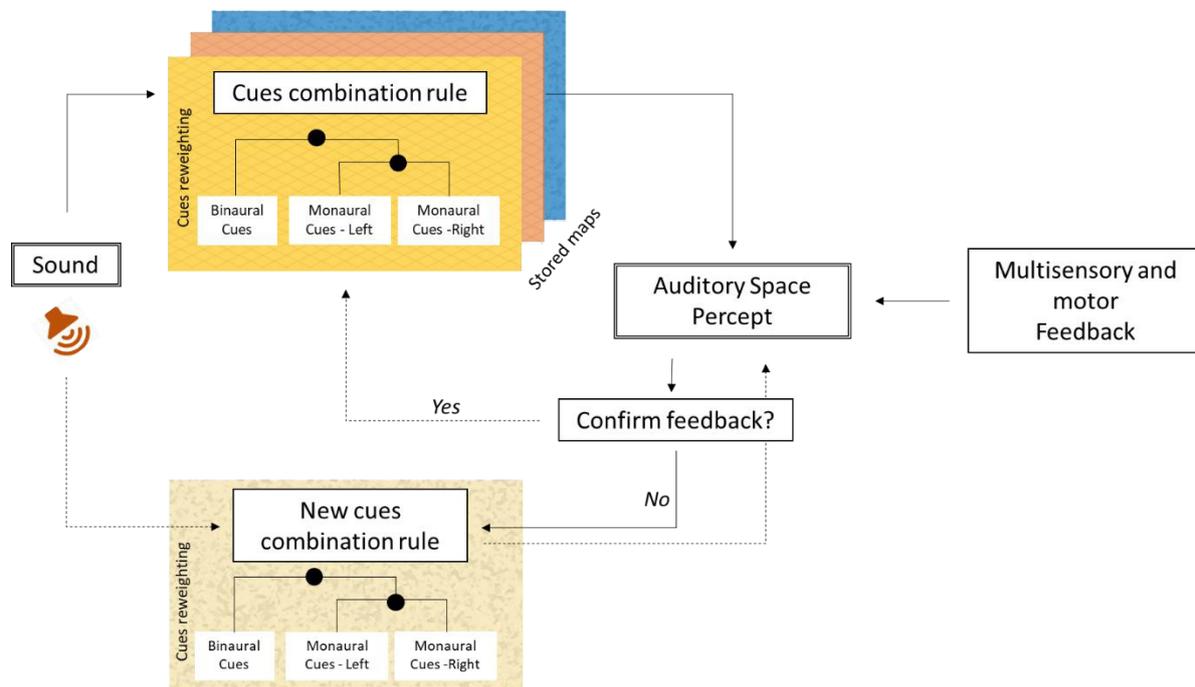


Figure 37 MODEL OF AUDITORY ADAPTATION PROCESSING.
The drawn was inspired and adapted from Mendonça, 2014.

7.2.1. Expanding the model: the contribution of active listening

In the following paragraph, I will discuss how the findings from the present thesis can further expand this model. In this thesis, I studied the importance of active listening and reaching-to-sound as a more naturalistic approach to sound localization, and as a training tool. This approach stems from the often-neglected notion that we are hardly static listeners in real life situations, and considers the fact that sound sources can also be reachable objects. In addition, it was inspired by previous works (Honda et al., 2007; Ohuchi et al., 2005; Parsehian & Katz, 2012; Valzolgher et al., 2020) in which acoustic space perception occurs in conditions that are more dynamic than the ones typically adopted in the lab. This approach can enrich the conceptual model in Figure 37, introducing the notion that motor interactions with sounds can be a continuous source of active and kinaesthetic feedback during sound localization.

In this active listening perspective, head-movements could play a prominent role. They can enrich the available auditory cues (Brimijoin & Akeroyd, 2012; Pastore, Natale, Yost, & Dorman, 2018; Wallach, 1940; Kato et al., 2003; Perrett & Noble, 1997a, 1997b), and they may contribute to identify the most appropriate cue combination rule for the current listening context. This was particularly evident throughout the experiments of this thesis. First, the ageing participants with presbycusis tested in Study 2 obtained a more veridical percept of auditory space when they were free to move their heads compared to when they

were static. Second, head movements played a key role when auditory perception was perturbed. In Study 3, I have shown a strong relationship between head-movements and changes in performance during the task. Furthermore, different head movements emerged as a function of task requirements (reaching vs naming). Finally, head-rotation behavior changed during training in normal hearing listener with one ear plug (Study 4) as well as CI users (Study 5).

When considering the contribution of active listening to the model, it is useful to draw a distinction between implicit and explicit learning mechanisms. This distinction has been introduced, for instance, in the framework developed by Huberdeau and colleagues (Huberdeau, Krakauer, & Haith, 2015; see also Taylor, Krakauer, & Ivry, 2014). They proposed to dissociate learning into two subcomponents: an implicit component, referring to a process of compensatory actions operating outside awareness, and an explicit component, referring to a process that operate within awareness and could represent a more definitive proof of the adaptation process that took place. Importantly, as suggested by Taylor and colleagues for the context of visuo-motor adaption studies, both explicit and implicit processes contribute to relearning and thus both need to be considered when studying its mechanisms.

Most previous studies on adaptation to novel auditory cues have usually focused on implicit relearning, by stressing the role of cue reweighting and cue remapping mechanisms. In this thesis, I consider instead the combined contribution of implicit and explicit learning components. Examples of implicit learning could be found, for instance, observed in the first head-movements after training (e.g., see Study 4). These movements occurred within the first second from sound onset and thus they may be interpreted as adaptive but operating outside awareness. By contrast, examples of explicit learning could be found in the progressive changes in head-movement exploration behavior that I observed during the reaching-to-sound task (see Study 3, 4 and 5). These movements may be considered as more strategic in nature, being implemented actively and consciously by the participants.

Expanding the focus to include both implicit and explicit learning mechanisms broadens also the consideration on the role of feedback during the adaptation to altered auditory cues. The feedback I have provided throughout the learning tasks allowed participants to compare their actual (and predicted) response with information about the actual position of the sound. This generated continuous instances for error detection. Feedback can trigger implicit learning by favoring cues-reweighting and cues-remapping, but it can also promote explicit learning by increasing participants awareness in two ways. First, though feedback, participants can notice the outcome of their performances and thus implement immediate actions aimed to reduce errors. Second, feedback gives them the possibility of testing these correcting actions and adjust them further to refine their adaptation strategies over time. Future works could clarify the contributions of feedback to implicit and explicit learning process.

Finally, considering implicit and explicit components together may shed light on the timing in which they contribute to the relearning process. Head movements implemented at the very beginning of the task could be considered as first spontaneous reactions to deal with a challenging hearing condition and thus they may be the result of an implicit process. Head movements observed later during the task, after a series of trials, may be implemented more consciously and may reflect an explicit process. In this respect, while implicit learning exists at an earlier stage of the learning process, explicit learning may require a longer time to be expressed and it may be the results of previous experience and successes. Being able to consciously extract information from feedback requires having experienced it several times. Future studies could explore explicit and implicit learning processes considering also this temporal aspect.

In sum, the distinction between implicit and explicit components may thus have useful implications for the understanding of head-movements during auditory space relearning. They can contribute to describe the complexity of the learning process exploiting the idea of dissociating it into sub-components (Huberdeau, Krakauer & Haith, 2015). At this stage, however, they are clearly speculations. Futures studies are needed to deepen more systematically the explicit and implicit contributions to learning in this field of research and to disentangle role of head behavioral strategies in the relearning process (see 7.4). Nonetheless, as concern head movements study, it was possible in this thesis to take a step forward as compared to previous literature because of the advantages provided by the VR methodological approach described in the next section (7.3).

7.3. Experimental work on listening require a paradigm shift

To study active listening, a change of methodological paradigm is required. This shift can be permitted by taking advantage of VR and motion tracking systems like the one used in the experimental contribution of the present doctoral work.

7.3.1. Measuring spatial hearing should occur in the three dimensions of space

The traditional approaches to study sound localization have constrained the most natural aspect of spatial hearing, that is the fact that auditory sources can occupy positions in the 3D space. The spatial position of sound has been rarely studied considering all the three dimensions of space (but see for notable pioneering exceptions: Brungart et al., 1999; Haber, Haber, Penningroth, Novak, & Radgowski, 1993). The reasons were likely related to technical limitations such as the restricted number of speakers available. Sounds typically were emitted from a limited set of positions, which varied along the azimuthal plane (Grantham et al., 2007), or changed both in azimuth and elevation from (Oldfield & Parker, 1984; Wightman & Kistler, 1999b). Furthermore, participants' response was often provided by egocentric method, as head or hand pointing, or using more allocentric methods in which participants reported the

sound position on a schematic representation of the environment (Bahu et al., 2016). These methods limit the possibility to measure the three dimensions of the perceived sound position. An interesting method to overcome this limitation is the one proposed in the pioneering work by Brungart and colleagues (1999) and presented in the study of Bahu and colleagues (2016) as “proximal pointing”. Participants indicated the perceived source direction by placing a marker held in their hand (Bahu, Carpentier, Noisternig, & Warusfel, 2016). Despite in this study the authors used this method only to detect the direction, in the discussion they emphasized the possibility of use this method also to assess the perceived distance of sound.

The pioneering work by Brungart and colleagues (Brungart et al., 1999) inspired the development of the system used in the experimental work of the present thesis, named *sphere* (European patent: 17723294.6-1115). Using sounds delivered at predefined head-referenced sound positions around the participant, together with real-time tracking of the listener’s hand, eyes and head, *sphere* allows a 3D cartography of acoustic space for each participant. This permits a precise assessment of auditory space processing. Sphere (Figure 38) allows (1) delivering real sounds at controlled and replicable positions; (2) alignment between the real auditory scene and virtual visual scenarios and full control of the visual stimuli presented (e.g. sound sources or scenes); and (3) real-time tracking of various body parts (notably head and hand) as well as eye-position monitoring. This allows full control over auditory and visual stimulation according to the experimental needs and, most importantly, it allows active listening. Participants’ head is tracked, and the stimuli can be delivered from head-centered starting position.

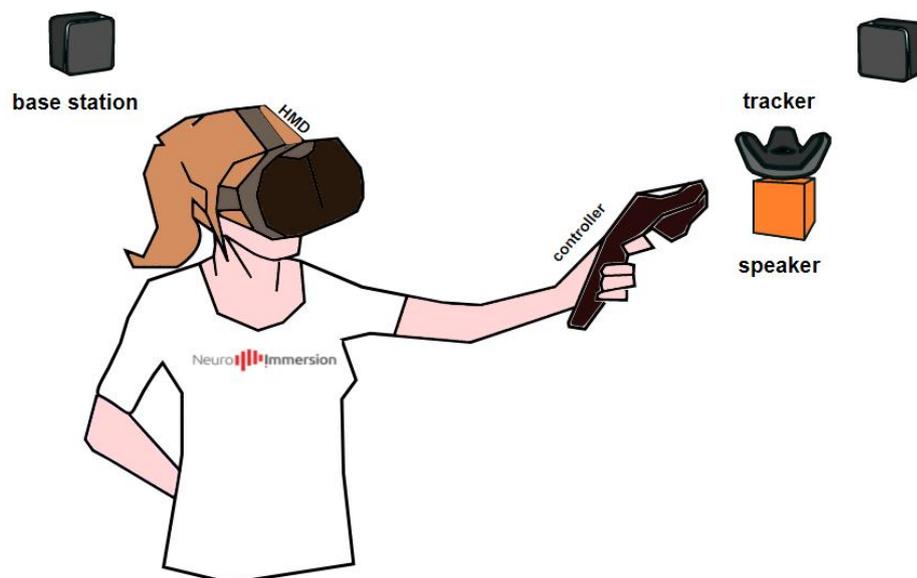


Figure 38 SCHEMATIC REPRESENTATION OF THE VR APPROCH

The two base stations permit the registration of the spatial position of HMD wearing by the participants, the tracker above the speaker and the controller headed by the participant.

This type of approach takes advantage of technology already available on the market. The HTC VIVE helmets used are produced and marketed starting from 2016 by Valve in collaboration with HTC. HTC Vive Systems comprises one head-mounted display (HMD), controllers, trackers (one mounted above the speaker to track its position in real time) and 2 lighthouse base stations (which serve for regular scanning of the position of controllers and trackers). Stimuli were controlled and delivered using a LDLC ZALMAN PC, using Steam VR software and the development platform Unity. This approach, further detailed in the methods sessions of each experimental work presented in this thesis, represents an innovative way to use virtual reality to study sound localization abilities without losing the advantage of using real free field sound.

7.3.2. Active listening should be allowed and measured

Despite the method proposed by Bahu (2016) gave to participants the possibility to be free to move the head, the authors instructed participants to remain still during listening. However, they tracked their head and measured some small involuntary head movements during the pointing phase. Despite the novelty of this method compared to the more traditional approaches to study sound localization, it constrained the natural exploration of the space through these limitations of head orienting. As reported in the previous session of the discussion, permitting head movements influences sound localization performance. In spite of this, more traditional studies have often constrained head movements to control the reproducibility of sound stimulation from consistent spatial positions using chin rest or asking participants to remain still (Brughera et al., 2013; Távora-Vieira, De Ceulaer, Govaerts, & Rajan, 2015). Recently, an interesting research conducted by Coudert and colleagues (2021) using the same VR technology proposed in this thesis tested the effect of a more natural hearing exploration during sound localization. The authors compared a static listening condition, in which bilateral cochlear implant children were asked to remain still during the sound presentation to a free head-motion condition, in which they were free to move the head as they want. They documented that spatial hearing abilities in bilateral cochlear implant children improve during free head-motion compared to head still condition. This finding represents important evidence of the potential of active listening and stresses the importance of having adequate technology to control and measure the complexity of spatial hearing. Despite this important contribution to the study of active listening, in their study head movements have not been measured systematically. Namely, head rotation and shift have not been quantified and studied. The studies proposed in this thesis have tried to overcome this limitation. Participants were tested letting them the possibility to move the head as they wanted (both during testing and training tasks). Measuring efficiently the kinematic of participants' head in a naturalistic condition has allowed to observe the spontaneous head strategies implemented by the participants, as well as their efficiency in improving sound localization performance.

7.3.3. Control of the visual scenarios and multisensory cues to sound position is essential

The multisensory contributions to sound localization have often been poorly controlled. Concerning vision, some studies allowed vision of the speakers, while others occluded them behind curtains (Nava, Bottari, Bonfioli, Beltrame, & Pavani, 2009; van Hoesel & Tyler, 2003) or blindfolded the participants (Ahrens, Lund, Marschall, & Dau, 2019). Both solutions hinder the possibility to control and manipulate the visual background and the direct visual information about source position independently. A method that overcomes this limit is the auditory virtual reality (Honda et al., 2007; Ohuchi et al., 2005; Parseihian & Katz, 2012; Sodnik, Tomazic, Grasset, Duenser, & Billinghamurst, 2006; see also Hobeika et al., 2020). Presenting spatialized sound through head-phone permits more flexibility in manipulating the visual scene. For instance, in the work of Steadman and colleagues (Steadman et al., 2019), auditory virtual reality was combined with visual virtual reality to obtain gamified training scenario in which participant were involved in game situations in which, for example, it was required to intercept sound sources by shooting visual and acoustic stimuli in motion. Being able to easily modify the visual scene is not only interesting to implement attractive training paradigms, but also allows to answer questions of a more theoretical nature. *Sphere* technology, for example, has made it possible to investigate the role of very basic visual information on the ability to locate sounds. As demonstrated in the Study 1 of this thesis, seeing an empty visual environment improves sound localization in elevation and produces faster eyes orienting responses to the sound. Thanks to sphere, it is possible to create different visual scenarios and thus to overcome the limitations of previous works. Sphere makes possible to investigate future research questions. For example, it would be interesting to manipulate the size of the virtual visual scene presented (without altering the real dimensions of the experiment room) to study the impact on sound localization abilities of the visual cues about the dimensions of the environment. Similarly, it could be possible to present environmental scenarios more or less rich in visual details.

7.3.4. A paradigm shift for experimental and clinical studies

However, in some cases, auditory virtual reality can limit the possibility to study people with hearing deficits. Despite it is possible to provide auditory stimuli directly connecting to the cochlear implant system, it remains unclear whether this type of stimulation can provide a comparable spatialization of sounds as compared to free field sounds. While auditory cues of normal hearing listeners have been studied and efficiently reproduced (Parseihian & Katz, 2012), for cochlear implant users the reliability of acoustic virtual reality is still an open-question. Furthermore, the necessity to let participants free to explore the space with the head requires further online computation to create reliable and coherent virtual sounds. Here too, sphere allows to overcome this limit, and, in this thesis, I tested both presbycusis and CI users using this

approach without encountering problems related to their hearing loss. Furthermore, its flexibility and usability, together with the affordable cost, make it suitable to be used in the future not only in a research context, but also as an integral part of clinical practice.

This technology, in addition to carrying out experimentally the idea of reaching sounds as a training tool and extracting precise indices of behavior, can also lead to image future experimental scenarios. Using this VR based approach is possible to create life-like simulations in the laboratory. The mechanisms underlying the behavior occurring in real life are not the same as those that occur in classical controlled experimental design. The main difference concerns the complexity of the natural context which is not fully replicable using classical methods (Shamay-Tsoory & Mendelsohn, 2019). Thus, to study the complexity of human behavior, cognitive sciences need to be able to create experimental paradigms closer to everyday life situations as suggested by Brunswik in 1949 with the term *ecological validity* (Brunswik, 1949). In this respect, virtual reality allows shortening this gap by proposing rich scenarios involving all the sensory modalities. Shamay-Tsoory and colleagues highlighted the importance of some key components which may improve the ecological validity of experimental settings. The first component is to allow the participant to be an active agent. The VR system adopted here allows to create listening context in which to act. Giving the possibility of moving is crucial and in line with the idea expressed by the construct of *embodied cognition*, which emphasizes the importance of bodily sensations and actions as part of cognition. Furthermore, as documented in this thesis, letting participants free to move can extend the study towards the strategic behaviors which can be adopted also in richer environments such as multi-talkers environments to deal with communication situation. The second component is to allow participants to be social agents. In this sense, especially when studying the understanding of the acoustic scene, introducing social stimuli may be fundamental. The acoustic scene is often a communicative situation in which one interacts with people and therefore the social aspect becomes relevant. It is important to consider the presence of other people also when investigating the implementation of active strategies. Some strategies to improve one's acoustic experience are related to interacting with other people. For example, asking someone to raise their hand before speaking or to increase the volume of their voice or to make their lips more visible are considered efficient behavioral strategies. Virtual reality technology allows to create these scenarios and future studies should take into account this further complexity and use technology to create appropriate naturalistic experiential settings.

7.4. Limitations and future directions

7.4.1. Clarify the role of behavioral strategies

In the present thesis behavioral strategies such as head movements emerged as a further element to be considered in the relearning process. To date, however, it is only possible to speculate about their role in

this process and how they can influence cue remapping and cue reweighting processes. Future studies are needed to disentangle the contribution of behavioral strategies from these other mechanisms subtending adaptation. They may be investigated systematically by creating experimental conditions in which the head-movements are impeded or permitted. In this way, it is possible to intermittently permit the implementation of behavioral strategies. A first hint in this direction has been presented in the discussion of the study 4. In this experiment, head movements were permitted during head pointing to sound, but they were limited using a chin-rest in the audio-visual task. However, in this study the differences in the two tasks prevented a precise control of this aspect. Whereas head pointing to sound is directly related to the acoustic space perception, audio-visual cueing required to localize sounds more indirectly. Alternating a steady head with an active head listening condition in a sound localization task that requires the same degree of ability to analyze the acoustic space would have allowed studying the contribution of the strategies in a more precise way. However, head-movements could also impact cue remapping and cue reweighting processes themselves. How can head-movements modulate these two processes? As already mentioned in the sessions 7.1. and 7.2., moving the head increases the available cues and the cues-reweighting mechanism could benefit from this: auditory cues experienced because of head-movements could also be considered and weighed to extract spatial coordinates of a certain sound source. Similarly, moving the head could modulate cues remapping. Accessing to new cues could facilitate the choice of a more suitable map among those that are already stored or building a new one. These hypotheses could be investigated in further studies permitting or preventing head-movements.

7.4.2. Including listening effort measures

A further element that may play an important role in the relearning process is the listening effort. While listening effort received considerable attention in the literature on hearing in noise (e.g., Klink, Schulte & Meis, 2012), it has been poorly investigated in the context of sound localization and in the experimental contributions of this thesis. However, it represents an important element that can influence improvement. In the here presented Study 3, I tried to do a step in this direction by asking participants to indicate how much effort, expressed in terms of energy and cognitive resources used, was required by the task. I found that listening effort increased during monaural listening as compared to the binaural one. However, I did not document changes across monaural block repetition. Higher listening effort was shown also in CI users as compared to normal hearing listeners both in a hearing in noise task and through a questionnaire investigating subjective measures of spatial hearing everyday experience as the Spatial Hearing Questionnaire (SHQ) (Perreau, Wu, Tatge, Irwin, & Corts, 2017). These findings revealed that CI users put more effort in their everyday listening experience as compared to normal hearing people. Measuring listening effort, either using self-rating questionnaires or through physiological indices (i.e. pupillometry, see for instance Visentin et al., 2021), can contribute to understand the subjective experience of

participants and to monitor the effect of certain conditions or trainings. In addition, higher perceived effort could be a decisive push in promoting the implementation of behavioral strategies, aimed not only at solving the task, but also at solving it with less effort.

This idea derives from the theoretical framework of cognitive offloading, the fundamental human ability to reduce cognitive demands and load for a given task through physical and social actions (Risko & Gilbert, 2016). An example of a behavior aiming at reducing effort is when people rotate their head to read rotated letters. This natural behavior, defined by Risko as an external normalization, represents a classical form of cognitive offloading. In other words, through body actions is possible to reduce the amount of effort required by a certain task (Risko, Medimorec, Chisholm, & Kingstone, 2014). In this regard, it is possible to interpret head movement actions as expressions of cognitive offloading because they aim at solving the task with less effort. This interpretation remains only a speculative: the limit of this thesis's studies is that they did not investigate effort as a dependent variable. Future studies could deepen this aspect further.

7.4.3. Expanding towards metacognitive belief

The implementation of strategies to solve the task does not derive only from the perception of effort but is also linked to the ability to judge one's own abilities, namely the metacognitive belief. Being able to judge correctly one's own behavior and identify difficulties and strengths means having good metacognition. Cognitive science identifies the metacognitive ability as one's own knowledge and cognition about cognitive phenomena (Flavell, 1979). Metacognition has been widely investigated in relation to learning, especially in the context of education (Azevedo & Alevin, 2013; Hartman, 2001; McCormick, 2003). Having effective metacognitive strategies is crucial to be a good learner (Tullis & Benjamin, 2011). Metacognitive skills can be studied by asking participants to rate their one performance and abilities. In this thesis only in study 3 I tried to take a step in this direction. Participants were asked to judge their performance from really bad to really good on a Likert scale. They were aware of their performance change after monaural plugging and of their improvement across monaural block repetition. Despite this evidence of participants' awareness of changes in their performance, a limitation of this thesis is not having tested it in the Study 4 and 5. It could have contributed to deepen the reaching to sound training paradigm's effect. Future studies may start to consider metacognitive abilities to both studying acoustic space perception and relearning and asking if and when the head-rotation strategy becomes 'aware' and purposefully applied.

A possible way to measure the participants' judgment about their performance is to ask to estimate the confidence about a certain answer provided. One interesting study, which has tried to deepen the relationship between confidence and performance in a sound localization task, is the work of Rabini and

colleagues (Rabini, Lucin, & Pavani, 2020). In this study, participants localised single sounds and were asked to provide a subjective rating of their spatial certainty about sound position. Whereas they were mostly accurate and certain about sound position in binaural listening, participants' accuracy and subjective certainty decreased in monaural listening. Intriguingly, during monaural condition accuracy and certainty dissociated. By studying performance and certainty across monaural trial repetitions, the authors revealed participants' certainty rapidly increased, but this did not happen for performance accuracy. These authors pointed out that a dissociation between accuracy and subjective certainty exists when mapping auditory input in the space. These observations may suggest that this dissociation may be a critical factor in the learning process.

7.4.4. Triggering efficient behavioral strategies

As demonstrated in the present thesis, leaving the freedom to explore, being active combined with multisensory feedback fostered acoustic space relearning. Reactions to multisensory feedback and behavioral strategies can also be considered as the results of more cognitive abilities of the participants. An example of such abilities is recognizing mindfully the effective and scarce strategies implemented. Another term related to the idea of knowing how to interpret correctly and regulate one's own behavior is self-regulation. Being aware of our difficulties while we are involved in an effortful cognitive task could promote the implementation of the necessary strategies to overcome or cope with them more easily and, hence, self-regulating ourselves. The important role of the metacognitive abilities in both learning and memory, and in regulating our own behavior is now well-established (Metcalfe and Finn, 2008; Nelson and Narens 1990). Being aware of the processes underlying behavior and performance, improves the ability to adapt to changing environments (Brigham & Pressley, 1988; Hertzog & Hultsch, 2000). In contrast, errors in metacognitive knowledge and monitoring can lead to deficient use of control while completing a task, decreasing performance (Karpicke, 2009; Tullis & Benjamin, 2011). The likelihood to implement environmental and behavioral adjustments is a direct consequence of how much people are aware of the difficulties they are experiencing, how much they can monitor and predict the outcome of their effective vs. ineffective behavior and how much they can update their knowledge about their skills to guide their behavior.

7.4.5. Thinking about internal feedback and confidence

All the training implemented in the thesis exploited feedback, as previous works suggested that receiving a feedback helps to adjust sound localization performance (Strelnikov et al., 2011). In our everyday life, we do not receive feedbacks for every action, but we rely nonetheless on comparisons between the expected consequences of our action and our actual performance. This may be considered a form of internal feedback. Are we able to detect an error using this feedback? Does this affect confidence about

our own performance? While the external feedback informs explicitly participants about their performance (likely impacting on their awareness) and the correct spatial position (providing precise spatial coordinate), internal feedback may operate in a different manner. For instance, it may impact on the sensation that the answer provided is correct or not, on the feeling that a certain task is easy or effortful. These internal elements could trigger adaptation to altered auditory conditions. Although I did not test it directly, I can speculate that a similar mechanism could have fostered relearning in the control condition proposed in Study 4 (Chapter 5), where participants were able to improve and modify the auditory space percept without relevant external feedback. Other studies documented spontaneous improvement (i.e. in control groups or conditions) which have been often attributed to habituation phenomena (Valzolgher et al., 2020; Rabini et al., 2019). Future studies may test the idea of internal feedback by measuring awareness and confidence of performance in the different testing situations.

7.4.6. Enriching the acoustic space perception model

The study of the metacognitive processes involved in relearning can therefore constitute an interesting future direction and it may enrich even further the model of auditory adaptation processing highlighting the continuous mutual influence between cues combination, strategies and metacognitive abilities. When the feedback challenges the percept, the entire combination of these elements needs to change: a new internal model awareness of one's own abilities needs to be created, new strategies could be implemented, and new cues combination rules must be tried. Furthermore, deepening the roles of cognitive effort, confidence and metacognitive elements lead to identify new ways to train people's acoustic abilities. Having an accurate representation of one's own abilities and working on implementing effective behavioral strategies to overcome difficulties can be crucial for patients. In this respect, future directions could concentrate on potentiating the metacognitive abilities of the patients. This may lead them to exploit more effectively the multisensory cues around them and induce them to try and test the effects of their own behavioral strategies.

In-depth information box
From lab to life: Tips and tools to hear better

This thesis work has contributed notions which may be useful for the clinical practice

Know the true potential of your patients!

Clinical settings are often not ecological. Testing patients with visible speakers can provide visual cues to sound position that may result in overestimation of sound localization abilities. However, asking patients to localize sound in a totally dark room is also not ecological, and brings the risk to underestimate their spatial hearing skills (see Study 1). Likewise, testing patients while limiting their possibility to listen actively to sounds may limit the use of the available auditory cues. As demonstrated in the Study 2 of this thesis, spontaneous head movements favour performance even in elderly participants with presbycusis.

Use all your multisensory resources!

Spatial hearing is neither unimodal, nor stationary. When localising sounds, patients can be encouraged to use also visual, somatosensory and motor cues to acquire information about the position of the auditory events. In the experimental contributions of this thesis, I have demonstrated that visual feedback, actions towards sounds, and exploration of the auditory environment with head movements can increase auditory spatial abilities in the ongoing task and they can also generalize to different experimental situations (Study 4). In this sense, encouraging patients to exploit the other sensory modalities and the active exploration of the environment may favour a more efficient analysis of the auditory scene. More in generally, it may help them in adjusting the correspondences between auditory cues and spatial coordinates.

Act! Use your body!

All the experimental evidence of the present thesis agrees in identifying head movements as crucial when locating sounds. They help to extract auditory cues and they are effective behavioural strategies when coping with altered listening conditions. Patients can approach the sources, turn the head, reach for the sound. The body is our first tool to perceive the world, patients could be incentivized to move and act in the surrounding space! Active behavioural strategies are not only useful for sound localization, but for speech intelligibility in both normal-hearing listeners (Grange & Culling, 2016) and in hearing-impaired listeners with asymmetric hearing loss (Brimijoin et al., 2012).

References

- Aarhus, R., Grönvall, E., Larsen, S. B., & Wollsen, S. (2011). Turning training into play: Embodied gaming, seniors, physical training and motivation. *Gerontechnology*, 10(2), 110–120. <https://doi.org/10.4017/gt.2011.10.2.005.00>
- Abel, S. M., & Hay, V. H. (1996). Sound localization the interaction of aging, hearing loss and hearing protection. *Scandinavian Audiology*, 25(1), 3-12.
- Abel, S. M., Giguère, C., Consoli, A., & Papsin, B. C. (2000). The effect of aging on horizontal plane sound localization. *The Journal of the Acoustical Society of America*, 108(2), 743–752. <https://doi.org/10.1121/1.429607>
- Agterberg, M. J. H., Snik, A. F. M., Hol, M. K. S., Van Wanrooij, M. M., & Van Opstal, A. J. (2012). Contribution of monaural and binaural cues to sound localization in listeners with acquired unilateral conductive hearing loss: Improved directional hearing with a bone-conduction device. *Hearing Research*, 286(1–2), 9–18. <https://doi.org/10.1016/j.heares.2012.02.012>
- Ahrens, A., Lund, K. D., Marschall, M., & Dau, T. (2019). Sound source localization with varying amount of visual information in virtual reality Axel. *PLoS ONE*, 14(3). <https://doi.org/10.1371/journal.pone.0214603>
- Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., ... Belliveau, J. W. (2006). Task-modulated “what” and “where” pathways in human auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 103(39), 14608–14613. <https://doi.org/10.1073/pnas.0510480103>
- Ahveninen, J., Kopčo, N., & Jääskeläinen, I. P. (2014). Psychophysics and neuronal bases of sound localization in humans. *Hearing Research*, Vol. 307, pp. 86–97. <https://doi.org/10.1016/j.heares.2013.07.008>
- Akeroyd, M. A. (2014). An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners. *Trends in Hearing*, 18, 233121651456044. <https://doi.org/10.1177/2331216514560442>
- Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). “What” and “where” in the human auditory system. *Proceedings of the National Academy of Sciences of the United States of America*, 98(21), 12301–12306. <https://doi.org/10.1073/pnas.211209098>
- Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, 14(3), 257–262. <https://doi.org/10.1016/j.cub.2004.01.029>
- Allen, M. C., Nikolopoulos, T. P., & O'Donoghue, G. M. (1998). Speech intelligibility in children after cochlear implantation. *The American journal of otology*, 19(6), 742-746.
- Angell, J. R., & Fite, W. (1901). From the Psychological Laboratory of the University of Chicago: The monaural localization of sound. *Psychological Review*, 8(3), 225–246. <https://doi.org/10.1037/h0073690>
- Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage*, 22(1), 401–408. <https://doi.org/10.1016/j.neuroimage.2004.01.014>

- Aronoff, J. M., Yoon, Y. S., Freed, D. J., Vermiglio, A. J., Pal, I., & Soli, S. D. (2010). The use of interaural time and level difference cues by bilateral cochlear implant users. *The Journal of the Acoustical Society of America*, 127(3), EL87-EL92.
- Asp, F., Mäki-Torkko, E., Karltop, E., Harder, H., Hergils, L., Eskilsson, G., & Stenfelt, S. (2012). Bilateral versus unilateral cochlear implants in children: *Speech recognition, sound localization, and parental reports*. *International journal of audiology*, 51(11), 817-832.
- Azevedo, R., & Alevén, V. (2013). Metacognition and learning technologies: An overview of current interdisciplinary research. *International handbook of metacognition and learning technologies*, 1-16. https://doi.org/10.1007/978-1-4419-5546-3_1
- Bahill, A. T., & McDonald, J. D. (1983). Frequency Limitations and Optimal Step Size for the Two-Point Central Difference Derivative Algorithm with Applications to Human Eye Movement Data. *IEEE Transactions on Biomedical Engineering*, BME-30(3), 191–194. <https://doi.org/10.1109/TBME.1983.325108>
- Bahu, H., Carpentier, T., Noisternig, M., & Warusfel, O. (2016). Comparison of different egocentric pointing methods for 3D sound localization experiments. *Acta Acustica United with Acustica*, 102(1), 107–118. <https://doi.org/10.3813/AAA.918928>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Bauer, R. W., Matuzsa, J. L., Blackmer, R. F., & Glucksberg, S. (1966). Noise Localization after Unilateral Attenuation. *The Journal of the Acoustical Society of America*, 40(2), 441–444. <https://doi.org/10.1121/1.1910093>
- Bear, M., Connors, B., & Paradiso, M. A. (2020). *Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain*. Jones & Bartlett Learning.
- Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin and Review*, 5(3), 482–489. <https://doi.org/10.3758/BF03208826>
- Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.
- Bolognini, N., Leor, F., Passamonti, C., Stein, B. E., & Làdavas, E. (2007). Multisensory-Mediated Auditory Localization. *Perception*, 36(10), 1477–1485. <https://doi.org/10.1068/p5846>
- Boyer, E. O., Babayan, B. M., Bevilacqua, F., Noisternig, M., Warusfel, O., Roby-Brami, A., ... Viaud-Delmon, I. (2013). From ear to hand: The role of the auditory-motor loop in pointing to an auditory source. *Frontiers in Computational Neuroscience*, (MAR). <https://doi.org/10.3389/fncom.2013.00026>
- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.
- Brimijoin, W. O., & Akeroyd, M. A. (2012). The Role of Head Movements and Signal Spectrum in an Auditory Front/Back Illusion. *I-Perception*, 3(3), 179–182. <https://doi.org/10.1068/i7173sas>
- Brimijoin, W. O., & Akeroyd, M. A. (2014). The moving minimum audible angle is smaller during self motion than during source motion. *Frontiers in neuroscience*, 8, 273.

- Brimijoin, W. O., & Akeroyd, M. A. (2016). The effects of hearing impairment, age, and hearing aids on the use of self-motion for determining front/back location. *Journal of the American Academy of Audiology*, 27(7), 588–600. <https://doi.org/10.3766/jaaa.15101>
- Brimijoin, W. O., McShefferty, D., & Akeroyd, M. A. (2010). Auditory and visual orienting responses in listeners with and without hearing-impairment. *The Journal of the Acoustical Society of America*, 127(6), 3678–3688. <https://doi.org/10.1121/1.3409488>
- Brimijoin, W. O., McShefferty, D., & Akeroyd, M. A. (2012). Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task. *Hearing Research*, 283(1–2), 162–168. <https://doi.org/10.1016/J.HEARES.2011.10.009>
- Brughera, A., Dunai, L., & Hartmann, W. M. (2013). Human interaural time difference thresholds for sine tones: The high-frequency limit. *The Journal of the Acoustical Society of America*, 133(5), 2839–2855. <https://doi.org/10.1121/1.4795778>
- Brunetti, M., Belardinelli, P., Caulo, M., Del Gratta, C., Della Penna, S., Ferretti, A., ... Romani, G. L. (2005). Human brain activation during passive listening to sounds from different locations: An fMRI and MEG study. *Human Brain Mapping*, 26(4), 251–261. <https://doi.org/10.1002/hbm.20164>
- Brungart, D. S., & Rabinowitz, W. M. (1999). Auditory localization of nearby sources. Head-related transfer functions. *The Journal of the Acoustical Society of America*, 106(3), 1465. <https://doi.org/10.1121/1.427180>
- Brungart, D. S., Durlach, N. I., & Rabinowitz, W. M. (1999). Auditory localization of nearby sources. II. Localization of a broadband source. *The Journal of the Acoustical Society of America*, 106(4), 1956–1968. <https://doi.org/10.1121/1.427943>
- Brunswik, E. (1949). Systematic and representative design of psychological experiments. With results in physical and social perception. In Proceedings of the [First] Berkeley symposium on mathematical statistics and probability (pp. 143-202). University of California Press.
- Bunn, A., & Korpela, M. (2019). An Introduction to dpIR. <https://doi.org/10.1016/j.dendro.2008.01.002>
- Calmels, M. N., Saliba, I., Wanna, G., Cochard, N., Fillaux, J., Deguine, O., & Fraysse, B. (2004). Speech perception and speech intelligibility in children after cochlear implantation. *International journal of pediatric otorhinolaryngology*, 68(3), 347-351.
- Carlile, S. (2014). The plastic ear and perceptual relearning in auditory spatial perception. *Frontiers in Neuroscience*, 8. <https://doi.org/10.3389/fnins.2014.00237>
- Carlile, S., Martin, R., & McAnally, K. (2005). Spectral Information in Sound Localization. *International Review of Neurobiology*, Vol. 70, pp. 399–434. [https://doi.org/10.1016/S0074-7742\(05\)70012-X](https://doi.org/10.1016/S0074-7742(05)70012-X)
- Chechik, G., Meilijson, I., & Ruppin, E. (1999). Neuronal regulation: A mechanism for synaptic pruning during brain maturation. *Neural Computation*, 11(8), 2061–2080. <https://doi.org/10.1162/089976699300016089>
- Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *Journal of the Acoustical Society of America*, 25(5), 975–979. <https://doi.org/10.1121/1.1907229>

- Clarke, S., Bellmann, A., Meuli, R. A., Assal, G., & Steck, A. J. (2000). Auditory agnosia and auditory spatial deficits following left hemispheric lesions: Evidence for distinct processing pathways. *Neuropsychologia*, 38(6), 797–807. [https://doi.org/10.1016/S0028-3932\(99\)00141-4](https://doi.org/10.1016/S0028-3932(99)00141-4)
- Clifton, R. K., Gwiazda, J., Bauer, J. A., Clarkson, M. G., & Held, R. M. (1988). Growth in Head Size During Infancy: Implications for Sound Localization. *Developmental Psychology*, 24(4), 477–483. <https://doi.org/10.1037/0012-1649.24.4.477>
- Clifton, R. K., Morrongiello, B. A., Kulig, J. W., & Dowd, J. M. (1981). Newborns' Orientation toward Sound: Possible Implications for Cortical Development. *Child Development*, 52(3), 833. <https://doi.org/10.2307/1129084>
- Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience*, Vol. 3, pp. 553–562. <https://doi.org/10.1038/nrn873>
- Colburn, H. S., & Kulkarni, A. (2005). Models of sound localization. In *Sound source localization* (pp. 272–316). Springer, New York, NY.
- Conti, S., Bonazzi, S., Laiacona, M., Masina, M., & Coralli, M. V. (2015). Montreal Cognitive Assessment (MoCA)-Italian version: regression based norms and equivalent scores. *Neurological Sciences*, 36(2), 209–214. <https://doi.org/10.1007/s10072-014-1921-3>
- Coudert, A., Gaveau, V., Gatel, J., Verdelet, G., Salemme, R., Farné, A., ... & Truy, E. (2021). Spatial Hearing Difficulties in Reaching Space in Bilateral Cochlear Implant Children Improve With Head Movements. *Ear and Hearing*.
- Cranford, J. L., Andres, M. A., Piatz, K. K., & Reissig, K. L. (1993). Influences of age and hearing loss on the precedence effect in sound localization. *Journal of Speech and Hearing Research*, 36(2), 437–441. <https://doi.org/10.1044/jshr.3602.437>
- Crespo, A. B., Idrovo, G. G., Rodrigues, N., & Pereira, A. (2016). A virtual reality UAV simulation with body area networks to promote the elders life quality. TISHW 2016 - 1st International Conference on Technology and Innovation in Sports, *Health and Wellbeing, Proceedings*. <https://doi.org/10.1109/TISHW.2016.7847780>
- Dehaene, S. (2018). *Apprendre!: les talents du cerveau, le défi des machines*. Odile Jacob.
- Denk, F., Ewert, S. D., & Kollmeier, B. (2019). On the limitations of sound localization with hearing devices. *The Journal of the Acoustical Society of America*, 146(3), 1732–1744. <https://doi.org/10.1121/1.5126521>
- Deouell, L. Y., Heller, A. S., Malach, R., D'Esposito, M., & Knight, R. T. (2007). Cerebral Responses to Change in Spatial Location of Unattended Sounds. *Neuron*, 55(6), 985–996. <https://doi.org/10.1016/j.neuron.2007.08.019>
- Dobrevá, M. S., O'Neill, W. E., & Paige, G. D. (2011). Influence of aging on human sound localization. *Journal of Neurophysiology*, 105(5), 2471–2486. <https://doi.org/10.1152/jn.00951.2010>
- Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, Vol. 10, pp. R731–R735. [https://doi.org/10.1016/S0960-9822\(00\)00740-5](https://doi.org/10.1016/S0960-9822(00)00740-5)

- Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America*, 76(1), 87-96.
- Eid, M., Gollwitzer, M. Schmitt, M., (2011). Statistik und Forschungsmethoden. Lehrbuch (2., korr. Aufl.) Grundlagen Psychologie;
- Eisenberg, L. S. (2007). Current state of knowledge: Speech recognition and production in children with hearing impairment. *Ear and Hearing*, 28(6), 766–772. <https://doi.org/10.1097/AUD.0B013E318157F01F>
- Etchemendy, P. E., Abregú, E., Calcagno, E. R., Eguia, M. C., Vechiatti, N., Iasi, F., & Vergara, R. O. (2017). Auditory environmental context affects visual distance perception. *Scientific Reports*, 7(1), 1–10. <https://doi.org/10.1038/s41598-017-06495-3>
- Field, J., Muir, D., Pilon, R., Sinclair, M., & Dodwell, P. (1980). Infants' Orientation to Lateral Sounds from Birth to Three Months. *Child Development*, 51(1), 295. <https://doi.org/10.2307/1129628>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, 34(10), 906.
- Forli, F., Lazzerini, F., Fortunato, S., Bruschini, L., & Berrettini, S. (2019). Cochlear implant in the elderly: results in terms of speech perception and quality of life. *Audiology and Neurotology*, 24(2), 77-83.
- Fox, J., & Weisberg, S. (2021). Using car and effects Functions in Other Functions. In cran.pau.edu.tr. Retrieved from <http://z.umn.edu/carbook>.
- Frissen, I., Vroomen, J., & de Gelder, B. (2012). The aftereffects of ventriloquism: the time course of the visual recalibration of auditory localization. *Seeing and perceiving*, 25(1), 1-14. <https://doi.org/10.1163/187847611X620883>
- Fu, Q. J., & Galvin III, J. J. (2007). Perceptual learning and auditory training in cochlear implant recipients. *Trends in Amplification*, 11(3), 193-205.
- Fuller, J. H. (1992). Head movement propensity. *Experimental Brain Research*, 92(1), 152-164.
- García-Betances, R. I., Jiménez-Mixco, V., Arredondo, M. T., & Cabrera-Umpiérrez, M. F. (2015). Using virtual reality for cognitive training of the elderly. *American Journal of Alzheimer's Disease and Other Dementias*, Vol. 30, pp. 49–54. <https://doi.org/10.1177/1533317514545866>
- Gates, G. A., Schmid, P., Kujawa, S. G., Nam, B. H., & D'Agostino, R. (2000). Longitudinal threshold changes in older men with audiometric notches. *Hearing Research*, 141(1–2), 220–228. [https://doi.org/10.1016/S0378-5955\(99\)00223-3](https://doi.org/10.1016/S0378-5955(99)00223-3)
- Gaveau, V., Coudert, A., Salemme, R., Koun, E., Desoche, C., Truy, E., ... Pavani, F. (2020). SPHERE: A novel approach to 3D and active sound localization. *BioRxiv*, 2020.03.19.998906. <https://doi.org/10.1101/2020.03.19.998906>
- Gibson, J. J. (1962). Observations on active touch. *Psychological Review*, 69(6), 477–491. <https://doi.org/10.1037/H0046962>
- Gifford, R. H., & Stecker, G. C. (2020). Binaural cue sensitivity in cochlear implant recipients with acoustic hearing preservation. *Hearing research*, 390, 107929.

- Gil-Carvajal, J. C., Cubick, J., Santurette, S., & Dau, T. (2016). Spatial Hearing with Incongruent Visual or Auditory Room Cues. *Scientific Reports*, 6(1), 1–10. <https://doi.org/10.1038/srep37342>
- Giovanelli, E., Valzolgher, C., Gessa, E., Todeschini, M., & Pavani, F. (2021). Unmasking the Difficulty of Listening to Talkers With Masks: lessons from the COVID-19 pandemic. *i-Perception*, 12(2), 204166952199839.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, Vol. 15, pp. 20–25. [https://doi.org/10.1016/0166-2236\(92\)90344-8](https://doi.org/10.1016/0166-2236(92)90344-8)
- Grange, J. A., & Culling, J. F. (2016). Head orientation benefit to speech intelligibility in noise for cochlear implant users and in realistic listening conditions. *The Journal of the Acoustical Society of America*, 140(6), 4061-4072.
- Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Labadie, R. F., & Haynes, D. S. (2007). Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants. *Ear and hearing*, 28(4), 524-541. <https://doi.org/10.1097/AUD.0b013e31806dc21a>
- Grieco-Calub, T. M., & Litovsky, R. Y. (2010). Sound localization skills in children who use bilateral cochlear implants and in children with normal acoustic hearing. *Ear and Hearing*, 31(5), 645–656. <https://doi.org/10.1097/AUD.0b013e3181e50a1d>
- Groh, J. M. (2014). Making space: how the brain knows where things are. Harvard University Press.
- Grothe, B., Pecka, M., & McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiological Reviews*, Vol. 90, pp. 983–1012. <https://doi.org/10.1152/physrev.00026.2009>
- Haber, L., Haber, R. N., Penningroth, S., Novak, K., & Radgowski, H. (1993). Comparison of Nine Methods of Indicating the Direction to Objects: Data from Blind Adults. *Perception*, 22(1), 35–47. <https://doi.org/10.1068/p220035>
- Hadley, L. V., Brimijoin, W. O., & Whitmer, W. M. (2019). Speech, movement, and gaze behaviours during dyadic conversation in noise. *Scientific Reports*, 9(1), 1–8. <https://doi.org/10.1038/s41598-019-46416-0>
- Haftner, E. R., & De Maio, J. (1975). Difference thresholds for interaural delay. *The Journal of the Acoustical Society of America*, 57(1), 181-187.
- Hartman, H. J. (Ed.). (2001). Metacognition in learning and instruction: Theory, research and practice (Vol. 19). *Springer Science & Business Media*. <https://doi.org/10.1007/978-94-017-2243-8>
- Heffner, H. E., & Heffner, R. S. (2014). The behavioral study of mammalian hearing. In *Perspectives on auditory research* (pp. 269-285). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-9102-6_15
- Heffner, H. E., & Heffner, R. S. (2018). The evolution of mammalian hearing. In *AIP Conference Proceedings* (Vol. 1965, No. 1, p. 130001). AIP Publishing LLC.
- Heffner, R. S. (1997). Comparative study of sound localization and its anatomical correlates in mammals. *Acta Oto-Laryngologica*, 117(sup532), 46-53.
- Heffner, R. S., & Heffner, H. E. (1992). Visual factors in sound localization in mammals. *The Journal of Comparative Neurology*, 317(3), 219–232. <https://doi.org/10.1002/cne.903170302>

- Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of comparative and physiological psychology*, 56(5), 872.
- Hendrickx, E., Stitt, P., Messonnier, J. C., Lyzwa, J. M., Katz, B. F., & de Boishéraud, C. (2017). Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. *The Journal of the Acoustical Society of America*, 141(3), 2011–2023. <https://doi.org/10.1121/1.4978612>
- Henning, D., Sabic, E., & Hout, M. C. (2018). Hear and There: Sounds from Everywhere! *Frontiers for Young Minds*, 6. <https://doi.org/10.3389/FRYM.2018.00063>
- Hertzog, C., & Hulstsch, D. F. (2000). Metacognition in adulthood and old age. igham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. *Psychology and aging*, 3(3), 249.
- Hobeika, L., Taffou, M., Carpentier, T., Warusfel, O., & Viaud-Delmon, I. (2020). Capturing the dynamics of peripersonal space by integrating expectancy effects and sound propagation properties. *Journal of neuroscience methods*, 332, 108534.
- Hofman, P. M., Van Riswick, J. G. A., & Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nature Neuroscience*, 1(5), 417–421. <https://doi.org/10.1038/1633>
- Honda, A., Shibata, H., Gyoba, J., Saitou, K., Iwaya, Y., & Suzuki, Y. (2007). Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game. *Applied Acoustics*, 68(8), 885–896. <https://doi.org/10.1016/j.apacoust.2006.08.007>
- Honda, A., Shibata, H., Hidaka, S., Gyoba, J., Iwaya, Y., & Suzuki, Y. (2013). Effects of head movement and proprioceptive feedback in training of sound localization. *I-Perception*, 4(4), 253–264. <https://doi.org/10.1068/i0522>
- Huberdeau, D. M., Krakauer, J. W., & Haith, A. M. (2015). Dual-process decomposition in human sensorimotor adaptation. *Current Opinion in Neurobiology*, 33, 71–77. <https://doi.org/10.1016/J.CONB.2015.03.003>
- Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, 50(3), 1202–1211. <https://doi.org/10.1016/j.neuroimage.2010.01.046>
- Irving, S., & Moore, D. R. (2011). Training sound localization in normal hearing listeners with and without a unilateral ear plug. *Hearing Research*, 280(1–2), 100–108. <https://doi.org/10.1016/j.heares.2011.04.020>
- Isaiah, A., & Hartley, D. E. (2015). Can training extend current guidelines for cochlear implant candidacy?. *Neural regeneration research*, 10(5), 718.
- Isaiah, A., Vongpaisal, T., King, A. J., & Hartley, D. E. H. (2014). Multisensory training improves auditory spatial processing following bilateral cochlear implantation. *Journal of Neuroscience*, 34(33), 11119–11130. <https://doi.org/10.1523/JNEUROSCI.4767-13.201>
- Iwaya, Y., Suzuki, Y., & Kimura, D. (2003). Effects of head movement on front-back error in sound localization. *Acoustical Science and Technology*, 24(5), 322–324. <https://doi.org/10.1250/ast.24.322>
- Johnen, A., Wagner, H., & Gaese, B. H. (2001). Spatial Attention Modulates Sound Localization in Barn Owls. *Journal of Neurophysiology*, 85(2), 1009–1012. <https://doi.org/10.1152/jn.2001.85.2.1009>

- Jones, H., Kan, A., & Litovsky, R. Y. (2014). Comparing sound localization deficits in bilateral cochlear-implant users and vocoder simulations with normal-hearing listeners. *Trends in Hearing*, 18, 233121651455457. <https://doi.org/10.1177/2331216514554574>
- Kacelnik, O., Nodal, F. R., Parsons, C. H., & King, A. J. (2006). Training-Induced Plasticity of Auditory Localization in Adult Mammals. *PLoS Biology*, 4(4), e71. <https://doi.org/10.1371/journal.pbio.0040071>
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138(4), 469.
- Kato, M., Uematsu, H., Kashino, M., & Hirahara, T. (2003). The effect of head motion on the accuracy of sound localization. *Acoustical Science and Technology*, Vol. 24, pp. 315–317. <https://doi.org/10.1250/ast.24.315>
- Keating, P., & King, A. J. (2013). Developmental plasticity of spatial hearing following asymmetric hearing loss: Context-dependent cue integration and its clinical implications. *Frontiers in Systems Neuroscience*, 7(DEC), 123. <https://doi.org/10.3389/fnsys.2013.00123>
- Keating, P., & King, A. J. (2015). Sound localization in a changing world. *Current Opinion in Neurobiology*, Vol. 35, pp. 35–43. <https://doi.org/10.1016/j.conb.2015.06.005>
- Keating, P., Dahmen, J. C., & King, A. J. (2013). Context-specific reweighting of auditory spatial cues following altered experience during development. *Current Biology*, 23(14), 1291–1299. <https://doi.org/10.1016/j.cub.2013.05.045>
- Keating, P., Dahmen, J. C., & King, A. J. (2015). Complementary adaptive processes contribute to the developmental plasticity of spatial hearing. *Nature Neuroscience*, 18(2), 185–187. <https://doi.org/10.1038/nn.3914>
- Keating, P., Rosenior-Patten, O., Dahmen, J. C., Bell, O., & King, A. J. (2016). Behavioral training promotes multiple adaptive processes following acute hearing loss. *ELife*, 5(MARCH2016). <https://doi.org/10.7554/eLife.12264>
- Kerber, S., & Seeber, B. U. (2012). Sound localization in noise by normal-hearing listeners and cochlear implant users. *Ear and Hearing*, 33(4), 445–457. <https://doi.org/10.1097/AUD.0b013e318257607b>
- Kim, J., Barnett-Cowan, M., & Macpherson, E. A. (2013). Integration of auditory input with vestibular and neck proprioceptive information in the interpretation of dynamic sound localization cues. *Proceedings of Meetings on Acoustics*, 19. <https://doi.org/10.1121/1.4799748>
- King A. J. (2004). The superior colliculus. *Current Biology*, 14(9), R335–R338.
- King, A. J., Hutchings, M. E., Moore, D. R., & Blakemore, C. (1988). Developmental plasticity in the visual and auditory representations in the mammalian superior colliculus. *Nature*, 332(6159), 73-76.
- King, A. J., Parsons, C. H., & Moore, D. R. (2000). Plasticity in the neural coding of auditory space in the mammalian brain. *Proceedings of the National Academy of Sciences*, 97(22), 11821–11828. <https://doi.org/10.1073/PNAS.97.22.11821>
- Klink, K. B., Schulte, M., & Meis, M. (2012). Measuring listening effort in the field of audiology—A literature review of methods (part 2). *Zeitschrift für Audiologie*, 51(2), 60-67

- Knudsen, E. I., & Knudsen, P. F. (1985). Vision guides the adjustment of auditory localization in young barn owls. *Science*, 230(4725), 545–548. <https://doi.org/10.1126/science.4048948>
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2009). Competition between auditory and visual spatial cues during visual task performance. *Experimental Brain Research*, 195(4), 593–602. <https://doi.org/10.1007/s00221-009-1829-y>
- Kokkinakis, K., Azimi, B., Hu, Y., & Friedland, D. R. (2012). Single and multiple microphone noise reduction strategies in cochlear implants. *Trends in amplification*, 16(2), 102-116. <https://doi.org/10.1177/1084713812456906>.
- Kolarik, A. J., Moore, B. C. J., Zahorik, P., Cirstea, S., & Pardhan, S. (2016). Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, and Psychophysics*, 78(2), 373–395. <https://doi.org/10.3758/s13414-015-1015-1>
- Kopinska, A., & Harris, L. R. (2003). Spatial representation in body coordinates: Evidence from errors in remembering positions of visual and auditory targets after active eye, head, and body movements. *Canadian Journal of Experimental Psychology*, 57(1), 23–37. <https://doi.org/10.1037/h0087410>
- Kral, A., Dorman, M. F., & Wilson, B. S. (2019). Neuronal development of hearing and language: cochlear implants and critical periods. *Annual review of neuroscience*, 42, 47-65. <https://doi.org/10.1146/annurev-neuro-080317-061513>
- Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, 12(4), 217-230. <https://doi.org/10.1038/nrn3008>
- Krogmann, R. J., & Al Khalili, Y. (2020). Cochlear Implants. StatPearls [Internet]. <https://www.ncbi.nlm.nih.gov/books/NBK544280/>
- Krumbholz, K., Schönwiesner, M., Von Cramon, D. Y., RübSamen, R., Shah, N. J., Zilles, K., & Fink, G. R. (2005). Representation of interaural temporal information from left and right auditory space in the human planum temporale and inferior parietal lobe. *Cerebral Cortex*, 15(3), 317–324. <https://doi.org/10.1093/cercor/bhh133>
- Kumpik, D. P., & King, A. J. (2019). A review of the effects of unilateral hearing loss on spatial hearing. *Hearing research*, 372, 17-28. <https://doi.org/10.1016/j.heares.2018.08.003>
- Kumpik, D. P., Kacelnik, O., & King, A. J. (2010). Adaptive reweighting of auditory localization cues in response to chronic unilateral earplugging in humans. *Journal of Neuroscience*, 30(14), 4883–4894. <https://doi.org/10.1523/JNEUROSCI.5488-09.2010>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13). <https://doi.org/10.18637/jss.v082.i13>
- Lederman, S. J., & Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, and Psychophysics*, Vol. 71, pp. 1439–1459. <https://doi.org/10.3758/APP.71.7.1439>
- Lewald, J., Foltys, H., & Töpper, R. (2002). Role of the posterior parietal cortex in spatial hearing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 22(3), RC207–RC207. <https://doi.org/10.1523/jneurosci.22-03-j0005.2002>

- Litovsky, R. Y., Ashmead, D. H., Gilkey, R., & Anderson, T. (1997). Development of binaural and spatial hearing in infants and children. *Binaural and spatial hearing in real and virtual environments*, 571-592.
- Litovsky, R. Y., Parkinson, A., & Arcaroli, J. (2009). Spatial hearing and speech intelligibility in bilateral cochlear implant users. *Ear and Hearing*, 30(4), 419-431. <https://doi.org/10.1097/AUD.0b013e3181a165be>
- Lomber, S. G., & Malhotra, S. (2008). Double dissociation of 'what' and 'where' processing in auditory cortex. *Nature Neuroscience*, 11(5), 609-616. <https://doi.org/10.1038/nn.2108>
- Lunner, T., Rudner, M., & Rönnerberg, J. (2009). Cognition and hearing aids. *Scandinavian journal of psychology*, 50(5), 395-403. <https://doi.org/10.1111/J.1467-9450.2009.00742.X>
- Luntz, M., Brodsky, A., Watad, W., Weiss, H., Tamir, A., & Pratt, H. (2005). Sound localization in patients with unilateral cochlear implants. *Cochlear implants international*, 6(1), 1-9.
- Macpherson, E. A. (2013). Cue weighting and vestibular mediation of temporal dynamics in sound localization via head rotation. *Proceedings of Meetings on Acoustics*, 19(1), 50131. <https://doi.org/10.1121/1.4799913>
- Macpherson, E. A., & Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, 111(5), 2219. <https://doi.org/10.1121/1.1471898>
- Majdak, P., Goupell, M. J., & Laback, B. (2010). 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. *Attention, Perception, and Psychophysics*, 72(2), 454-469. <https://doi.org/10.3758/APP.72.2.454>
- Majdak, P., Walder, T., & Laback, B. (2013). Effect of long-term training on sound localization performance with spectrally warped and band-limited head-related transfer functions. *The Journal of the Acoustical Society of America*, 134(3), 2148-2159. <https://doi.org/10.1121/1.4816543>
- Makous, J. C., & Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *Journal of the Acoustical Society of America*, 87(5), 2188-2200. <https://doi.org/10.1121/1.399186>
- McAnally, K. I., & Martin, R. L. (2014). Sound localization with head movement: implications for 3-d audio displays. *Frontiers in Neuroscience*, 8(8 JUL), 210. <https://doi.org/10.3389/fnins.2014.00210>
- McCormick, C. B. (2003). Metacognition and Learning. In *Handbook of Psychology* (pp. 79-102). <https://doi.org/10.1002/0471264385.wei0705>
- Mendonça, C. (2014). A review on auditory space adaptations to altered head-related cues. *Frontiers in Neuroscience*, Vol. 8. <https://doi.org/10.3389/fnins.2014.00219>
- Mendonça, C., Campos, G., Dias, P., & Santos, J. A. (2013). Learning Auditory Space: Generalization and Long-Term Effects. *PLoS ONE*, 8(10), e77900. <https://doi.org/10.1371/journal.pone.0077900>
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15(1), 174-179.
- Middlebrooks, J. C. (2015). Sound localization. In *Handbook of Clinical Neurology* (Vol. 129, pp. 99-116). <https://doi.org/10.1016/B978-0-444-62630-1.00006-8>

- Middlebrooks, J. C., & Green, D. M. (1991). Sound Localization by Human Listeners. *Annual Review of Psychology*, 42(1), 135–159. <https://doi.org/10.1146/annurev.ps.42.020191.001031>
- Middlebrooks, J., Simon, J., Popper, A., & Fay, R. (2017). The Auditory System at the Cocktail Party. In *The Auditory System at the Cocktail Party*. <https://doi.org/10.1007/978-3-319-51662-2>
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1), 20-25.
- Molina, K. I., Ricci, N. A., De Moraes, S. A., & Perracini, M. R. (2014). Virtual reality using games for improving physical functioning in older adults: A systematic review. *Journal of NeuroEngineering and Rehabilitation*, Vol. 11, pp. 1–20. <https://doi.org/10.1186/1743-0003-11-156>
- Moore, D. R., & Shannon, R. V. (2009). Beyond cochlear implants: Awakening the deafened brain. *Nature Neuroscience*, Vol. 12, pp. 686–691. <https://doi.org/10.1038/nn.2326>
- Morikawa, D., & Hirahara, T. (2013). Effect of head rotation on horizontal and median sound localization of band-limited noise. *Acoustical Science and Technology*, 34(1), 56–58. <https://doi.org/10.1250/ast.34.56>
- Morikawa, D., Toyoda, Y., & Hirahara, T. (2011). Impact of head movement on sound localization with band-limited noise. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (Vol. 2011, No. 6, pp. 1963-1968). Institute of Noise Control Engineering.
- Morikawa, D., Toyoda, Y., & Hirahara, T. (2013). Head movement during horizontal and median sound localization experiments in which head-rotation is allowed. *Proceedings of Meetings on Acoustics*, 19. <https://doi.org/10.1121/1.4800148>
- Morrongiello, B. A., & Rocca, P. T. (1987). Infants' Localization of Sounds in the Horizontal Plane: Effects of Auditory and Visual Cues. *Child Development*, 58(4), 918. <https://doi.org/10.2307/1130532>
- Moua, K., Kan, A., Jones, H. G., Misurelli, S. M., & Litovsky, R. Y. (2019). Auditory motion tracking ability of adults with normal hearing and with bilateral cochlear implants. *The Journal of the Acoustical Society of America*, 145(4), 2498–2511. <https://doi.org/10.1121/1.5094775>
- Muir, D. W., Clifton, R. K., & Clarkson, M. G. (1989). The development of a human auditory localization response: a U-shaped function. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 43(2), 199. [10.1037/h0084220](https://doi.org/10.1037/h0084220)
- Muir, D. W., Humphrey, D. E., & Humphrey, G. K. (1994). Pattern and space perception in young infants. *Spatial Vision*.
- Muir, D., & Field, J. (1979). Newborn Infants Orient to Sounds. *Child Development*, 50(2), 431. <https://doi.org/10.2307/1129419>
- Musicant, A. D., & Butler, R. A. (1980). Monaural localization: An analysis of practice effects. *Perception & Psychophysics*, 28(3), 236–240. <https://doi.org/10.3758/BF03204379>
- Nava, E., Bottari, D., Bonfioli, F., Beltrame, M. A., & Pavani, F. (2009). Spatial hearing with a single cochlear implant in late-implanted adults. *Hearing Research*, 255(1–2), 91–98. <https://doi.org/10.1016/j.heares.2009.06.007>

- Nawaz, S., McNeill, C., & Greenberg, S. L. (2014). Improving Sound Localization After Cochlear Implantation and Auditory Training for the Management of Single-Sided Deafness. *Otology & Neurotology*, 35(2), 271–276. <https://doi.org/10.1097/MAO.0000000000000257>
- Neher, T., Laugesen, S., Søgaaard Jensen, N., & Kragelund, L. (2011). Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities? *The Journal of the Acoustical Society of America*, 130(3), 1542–1558. <https://doi.org/10.1121/1.3608122>
- Németh, R., Háden, G. P., Török, M., & Winkler, I. (2015). Processing of horizontal sound localization cues in newborn infants. *Ear and hearing*, 36(5), 550-556. 10.1097/AUD.0000000000000160
- Noble W.G. (1981) Earmuffs, exploratory head movements, and horizontal and vertical sound localization. *Journal of Auditory Research*. 21,1–12.
- Noble, W., Byrne, D., & Lepage, B. (1994). Effects on sound localization of configuration and type of hearing impairment. *Journal of the Acoustical Society of America*, 95(2), 992–1005. <https://doi.org/10.1121/1.408404>
- Nothwang, H. G. (2016). Evolution of mammalian sound localization circuits: A developmental perspective. *Progress in Neurobiology*, 141, 1–24. <https://doi.org/10.1016/J.PNEUROBIO.2016.02.003>
- Noyce, A. L., Cestero, N., Michalka, S. W., Shinn-Cunningham, B. G., & Somers, D. C. (2017). Sensory-biased and multiple-demand processing in human lateral frontal cortex. *Journal of Neuroscience*, 37(36), 8755–8766. <https://doi.org/10.1523/JNEUROSCI.0660-17.2017>
- Oberfeld, D., & Kloeckner-Nowotny, F. (2016). Individual differences in selective attention predict speech identification at a cocktail party. *Elife*, 5, e16747. <https://doi.org/10.7554/eLife.16747>
- Ohuchi, M., Iwaya, Y., Suzuki, Y., & Munekata, T. (2005). Training effect of a virtual auditory game on sound localization ability of the visually impaired. In: *Proceedings of ICAD 2005*, Limerick, Ireland; 2005, July 6–9. p. 283.
- Oldfield, S. R., & Parker, S. P. (1984). Acuity of sound localization: a topography of auditory space. I. Normal hearing conditions. *Perception*, 13(5), 581–600. <https://doi.org/10.1068/p130581>
- Orzan, E., Rota, E., Crovetti, F., De Colle, W. Piccola Guida per i genitori di bambini con Problemi di Udito, retrieved from: <https://www.burlo.trieste.it/audiologia-ori>
- Otte, R. J., Agterberg, M. J. H., Van Wanrooij, M. M., Snik, A. F. M., & Van Opstal, A. J. (2013). Age-related hearing loss and ear morphology affect vertical but not horizontal sound-localization performance. *JARO - Journal of the Association for Research in Otolaryngology*, 14(2), 261–273. <https://doi.org/10.1007/s10162-012-0367-7>
- Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences of the United States of America*, 111(16), 6104–6108. <https://doi.org/10.1073/pnas.1322705111>
- Parseihian, & Katz. (2012). Rapid head-related transfer function adaptation using a virtual auditory environment. *The Journal of the Acoustical Society of America*, 131(4), 2948–2957. <https://doi.org/10.1121/1.3687448>

- Parseihian, Jouffrais, & Katz, B. F. G. (2014). Reaching nearby sources: comparison between real and virtual sound and visual targets. *Frontiers in Neuroscience*, 8(SEP), 269. <https://doi.org/10.3389/fnins.2014.00269>
- Pastore, M. T., Natale, S., Yost, W., & Dorman, M. F. (2018). Head movements allow listeners bilaterally implanted with cochlear implants to resolve front-back confusions. *Ear and hearing*, 39(6), 1224. <https://doi.org/10.1097/AUD.0000000000000581>
- Pavani, F., Farnè, A., & Làdavas, E. (2003). Task-dependent visual coding of sound position in visuospatial neglect patients. *Neuroreport*, 14(1), 99-103.
- Pavani, F., Husain, M., & Driver, J. (2008). Eye-movements intervening between two successive sounds disrupt comparisons of auditory location. *Experimental brain research*, 189(4), 435-449. <https://doi.org/10.1007/s00221-008-1440-7>
- Pavani, F., Meneghello, F., & Làdavas, E. (2001). Deficit of auditory space perception in patients with visuospatial neglect. *Neuropsychologia*, 39(13), 1401–1409. [https://doi.org/10.1016/S0028-3932\(01\)00060-4](https://doi.org/10.1016/S0028-3932(01)00060-4)
- Pavani, F., Venturini, M., Baruffaldi, F., Artesini, L., Bonfioli, F., Frau, G. N., & van Zoest, W. (2017). Spatial and non-spatial multisensory cueing in unilateral cochlear implant users. *Hearing Research*, 344, 24–37. <https://doi.org/10.1016/j.heares.2016.10.025>
- Pearson, J. D., Morrell, C. H., Gordon-Salant, S., Brant, L. J., Jeffrey Metter, E., Klein, L. L., & Fozard, J. L. (1995). Gender differences in a longitudinal study of age-associated hearing loss. *Journal of the Acoustical Society of America*, 97(2), 1196–1205. <https://doi.org/10.1121/1.412231>
- Perreau, A. E., Wu, Y. H., Tatge, B., Irwin, D., & Corts, D. (2017). Listening effort measured in adults with normal hearing and cochlear implants. *Journal of the American Academy of Audiology*, 28(8), 685–697. <https://doi.org/10.3766/JAAA.16014/BIB>
- Perrett, S., & Noble, W. (1997a). The contribution of head motion cues to localization of low-pass noise. *Perception and Psychophysics*, 59(7), 1018–1026. <https://doi.org/10.3758/BF03205517>
- Perrett, S., & Noble, W. (1997b). The effect of head rotations on vertical plane sound localization. *The Journal of the Acoustical Society of America*, 102(4), 2325–2332. <https://doi.org/10.1121/1.419642>
- Plancher, G., Gyselinck, V., Nicolas, S., & Piolino, P. (2010). Age effect on components of episodic memory and feature binding: A virtual reality study. *Neuropsychology*, 24(3), 379–390. <https://doi.org/10.1037/a0018680>
- Poirier, C., Baumann, S., Dheerendra, P., Joly, O., Hunter, D., Balezeau, F., ... Griffiths, T. D. (2017). Auditory motion-specific mechanisms in the primate brain. *PLOS Biology*, 15(5), e2001379. <https://doi.org/10.1371/journal.pbio.2001379>
- Poirier-Quinot, D., & Katz, B. (2020). Assessing the Impact of Head-Related Transfer Function Individualization on Task Performance: Case of a Virtual Reality Shooter Game. *Journal of the Audio Engineering Society*, 68(4), 248–260. <https://doi.org/10.17743/jaes.2020.0004>
- Poirier-Quinot, D., & Katz, B. F. (2018). Impact of HRTF individualization on player performance in a VR shooter game II. In *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society.

- Pollack, I., & Rose, M. (1967). Effect of head movement on the localization of sounds in the equatorial plane. *Perception & Psychophysics*, 2(12), 591–596. <https://doi.org/10.3758/BF03210274>
- Pöntynen, H., & Salminen, N. H. (2019). Resolving front-back ambiguity with head rotation: The role of level dynamics. *Hearing Research*, 377, 196–207. <https://doi.org/10.1016/j.heares.2019.03.020>
- Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1), 3–25. <https://doi.org/10.1080/00335558008248231>
- Pumphrey, R. J. (1950). Hearing. In: *Physiological Mechanisms in Animal Behavior*. Cambridge, U.K. *Symposium of the Society for Experimental Biology* (Vol. 4, pp. 3–18).
- Rabini, G., Altobelli, E., & Pavani, F. (2019). Interactions between egocentric and allocentric spatial coding of sounds revealed by a multisensory learning paradigm. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-44267-3>
- Rakerd, B., Vander Velde, T. J., & Hartmann, W. M. (1998). Sound localization in the median sagittal plane by listeners with presbycusis. *Journal of the American Academy of Audiology*, 9(6), 466–479.
- Rauschecker, J. P. (2018). Where, When, and How: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex*, Vol. 98, pp. 262–268. <https://doi.org/10.1016/j.cortex.2017.10.020>
- Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 97, pp. 11800–11806. <https://doi.org/10.1073/pnas.97.22.11800>
- Rayes, H., Al-Malky, G., & Vickers, D. (2019). Systematic review of auditory training in pediatric cochlear implant recipients. *Journal of Speech, Language, and Hearing Research*, 62(5), 1574–1593.
- Rayleigh, Lord. (1907). XII. On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(74), 214–232. <https://doi.org/10.1080/14786440709463595>
- Redon, C., & Hay, L. (2005). Role of visual context and oculomotor conditions in pointing accuracy. *Neuroreport*, 16(18), 2065–2067.
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, Vol. 20, pp. 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>
- Risko, E. F., Medimorec, S., Chisholm, J., & Kingstone, A. (2014). Rotating With Rotated Text: A Natural Behavior Approach to Investigating Cognitive Offloading. *Cognitive Science*, 38(3), 537–564. <https://doi.org/10.1111/cogs.12087>
- Risoud, M., Hanson, J. N., Gauvrit, F., Renard, C., Lemesre, P. E., Bonne, N. X., & Vincent, C. (2018). Sound source localization. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, Vol. 135, pp. 259–264. <https://doi.org/10.1016/j.anorl.2018.04.009>
- Saenz, M., & Langers, D. R. M. (2014). Tonotopic mapping of human auditory cortex. *Hearing Research*, Vol. 307, pp. 42–52. <https://doi.org/10.1016/j.heares.2013.07.016>

- Schubotz, R. I., Von Cramon, D. Y., & Lohmann, G. (2003). Auditory what, where, and when: A sensory somatotopy in lateral premotor cortex. *NeuroImage*, 20(1), 173–185. [https://doi.org/10.1016/S1053-8119\(03\)00218-0](https://doi.org/10.1016/S1053-8119(03)00218-0)
- Schutte, M., Ewert, S. D., & Wiegrefe, L. (2019). The percept of reverberation is not affected by visual room impression in virtual environments. *The Journal of the Acoustical Society of America*, 145(3), EL229–EL235. <https://doi.org/10.1121/1.5093642>
- Seeber, B. U., & Fastl, H. (2008). Localization cues with bilateral cochlear implants. *The Journal of the Acoustical Society of America*, 123(2), 1030-1042.
- Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-life neuroscience: an ecological approach to brain and behavior research. *Perspectives on Psychological Science*, 14(5), 841-859. <https://doi.org/10.1177/1745691619856350>
- Sharma, A., Dorman, M. F., & Spahr, A. J. (2002). A Sensitive Period for the Development of the Central Auditory System in Children with Cochlear Implants: Implications for Age of Implantation. *Ear & Hearing*, 23, 532–539. <https://doi.org/10.1097/01.AUD.0000042223.62381.01>
- Shelton, B. R., & Searle, C. L. (1980). The influence of vision on the absolute identification of sound-source position. *Perception & Psychophysics*, 28(6), 589–596. <https://doi.org/10.3758/BF03198830>
- Shepherd, R. K., Seligman, P. M., & Fallon, J. B. (2015). Neuroprostheses for restoring hearing loss. *Implantable Neuroprostheses for Restoring Function*, 97–125. <https://doi.org/10.1016/B978-1-78242-101-6.00005-7>
- Shinn-Cunningham, B. G., Durlach, N. I., & Held, R. M. (1998). Adapting to supernormal auditory localization cues. I. Bias and resolution. *The Journal of the Acoustical Society of America*, 103(6), 3656. <https://doi.org/10.1121/1.423088>
- Shinn-Cunningham, B. G., Santarelli, S., & Kopco, N. (2000). Tori of confusion: Binaural localization cues for sources within reach of a listener. *The Journal of the Acoustical Society of America*, 107(3), 1627–1636. <https://doi.org/10.1121/1.428447>
- Shinn-Cunningham, B., Best, V., & Lee, A. K. (2017). Auditory object formation and selection. In *The auditory system at the cocktail party* (pp. 7-40). Springer, Cham. https://doi.org/10.1007/978-3-319-51662-2_2
- Smulders, Y. E., van Zon, A., Stegeman, I., Rinia, A. B., Van Zanten, G. A., Stokroos, R. J., ... & Grolman, W. (2016). Comparison of bilateral and unilateral cochlear implantation in adults: a randomized clinical trial. *JAMA otolaryngology–head & neck surgery*, 142(3), 249-256.
- Sodnik, J., Tomazic, S., Grasset, R., Duenser, A., & Billinghamurst, M. (2006). Spatial sound localization in an augmented reality environment. *ACM International Conference Proceeding Series*, 206, 111–118. <https://doi.org/10.1145/1228175.1228197>
- Spence, C., & Driver, J. (1996). Audiovisual Links in Endogenous Covert Spatial Attention. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4), 1005–1030. <https://doi.org/10.1037/0096-1523.22.4.1005>

- Spierer, L., Bellmann-Thiran, A., Maeder, P., Murray, M. M., & Clarke, S. (2009). Hemispheric competence for auditory spatial representation. *Brain*, 132(7), 1953–1966. <https://doi.org/10.1093/brain/awp127>
- Steadman, M. A., Kim, C., Lestang, J. H., Goodman, D. F. M., & Picinali, L. (2019). Short-term effects of sound localization training in virtual reality. *Scientific Reports*, 9(1), 1–17. <https://doi.org/10.1038/s41598-019-54811-w>
- Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America*, 116(2), 1081-1091.
- Stitt, P., Picinali, L., & Katz, B. F. G. (2019). Auditory Accommodation to Poorly Matched Non-Individual Spectral Localization Cues Through Active Learning. *Scientific Reports*, 9(1), 1–14. <https://doi.org/10.1038/s41598-018-37873-0>
- Strelnikov, K., Rosito, M., & Barone, P. (2011). Effect of Audiovisual Training on Monaural Spatial Hearing in Horizontal Plane. *PLoS ONE*, 6(3), e18344. <https://doi.org/10.1371/journal.pone.0018344>
- Tardif, E., Murray, M. M., Meylan, R., Spierer, L., & Clarke, S. (2006). The spatio-temporal brain dynamics of processing and integrating sound localization cues in humans. *Brain Research*, 1092(1), 161–176. <https://doi.org/10.1016/j.brainres.2006.03.095>
- Tasell, D. J. V. (1993). Hearing loss, speech, and hearing aids. *Journal of Speech, Language, and Hearing Research*, 36(2), 228-244.
- Távora-Vieira, D., De Ceulaer, G., Govaerts, P. J., & Rajan, G. P. (2015). Cochlear Implantation Improves Localization Ability in Patients With Unilateral Deafness. *Ear and Hearing*, 36(3), e93–e98. <https://doi.org/10.1097/AUD.0000000000000130>
- Taylor, J. A., Krakauer, J. W., & Ivry, R. B. (2014). Explicit and Implicit Contributions to Learning in a Sensorimotor Adaptation Task. *Journal of Neuroscience*, 34(8), 3023–3032. <https://doi.org/10.1523/JNEUROSCI.3619-13.2014>
- Thurlow, W. R., & Runge, P. S. (1967). Effect of Induced Head Movements on Localization of Direction of Sounds. *The Journal of the Acoustical Society of America*, 42(2), 480–488. <https://doi.org/10.1121/1.1910604>
- Tonelli, A., Brayda, L., & Gori, M. (2015). Task-dependent calibration of auditory spatial perception through environmental visual observation. *Frontiers in Systems Neuroscience*, 9(June), 84. <https://doi.org/10.3389/fnsys.2015.00084>
- Trapeau, R., & Schönwiesner, M. (2015). Adaptation to shifted interaural time differences changes encoding of sound location in human auditory cortex. *NeuroImage*, 118, 26–38. <https://doi.org/10.1016/j.neuroimage.2015.06.006>
- Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language*, 64(2), 109–118. <https://doi.org/10.1016/j.jml.2010.11.002>
- Tyler, R. S., Witt, S. A., Dunn, C. C., & Wang, W. (2010). Initial development of a spatially separated speech-in-noise and localization training program. *Journal of the American Academy of Audiology*, 21(06), 390-403.

- Valzolgher, C., Alzhaler, M., Gessa, E., Todeschini, M., Nieto, P., Verdelet, G., ... Pavani, F. (2020). The impact of a visual spatial frame on real sound-source localization in virtual reality. *Current Research in Behavioral Sciences*, 1, 100003. <https://doi.org/10.1016/j.crbeha.2020.100003>
- Valzolgher, C., Campus, C., Rabini, G., Gori, M., & Pavani, F. (2020). Updating spatial hearing abilities through multisensory and motor cues. *Cognition*, 204, 104409. <https://doi.org/10.1016/j.cognition.2020.104409>
- Valzolgher, C., Verdelet, G., Salemme, R., Lombardi, L., Gaveau, V., Farné, A., & Pavani, F. (2020). Reaching to sounds in virtual reality: A multisensory-motor approach to promote adaptation to altered auditory cues. *Neuropsychologia*, 149, 107665. <https://doi.org/10.1016/j.neuropsychologia.2020.107665>
- Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., & Wouters, J. (2006). Horizontal localization with bilateral hearing aids: Without is better than with. *The Journal of the Acoustical Society of America*, 119(1), 515–526. <https://doi.org/10.1121/1.2139653>
- van der Heijden, K., Rauschecker, J. P., de Gelder, B., & Formisano, E. (2019). Cortical mechanisms of spatial hearing. *Nature Reviews Neuroscience*, 20(10), 609–623. Retrieved from <https://www.nature.com/articles/s41583-019-0206-5>
- van der Heijden, Kiki, Rauschecker, J. P., Formisano, E., Valente, G., & de Gelder, B. (2018). Active sound localization sharpens spatial tuning in human primary auditory cortex. *Journal of Neuroscience*, 38(40), 8574–8587. <https://doi.org/10.1523/JNEUROSCI.0587-18.2018>
- van der Meer, A. L., Ramstad, M., & van der Weel, F. R. (2008). Choosing the shortest way to mum: Auditory guided rotation in 6-to 9-month-old infants. *Infant Behavior and Development*, 31(2), 207-216. <https://doi.org/10.1016/j.infbeh.2007.10.007>
- Van der Zwaag, W., Gentile, G., Gruetter, R., Spierer, L., & Clarke, S. (2011). Where sound position influences sound object representations: A 7-T fMRI study. *NeuroImage*, 54(3), 1803–1811. <https://doi.org/10.1016/j.neuroimage.2010.10.032>
- van Hoesel, R. J. M., & Tyler, R. S. (2003). Speech perception, localization, and lateralization with bilateral cochlear implants. *The Journal of the Acoustical Society of America*, 113(3), 1617–1630. <https://doi.org/10.1121/1.1539520>
- Van Opstal, J. (2016). *The auditory system and human sound-localization behavior*. Academic Press.
- Van Wanrooij, M. M., & Van Opstal, A. J. (2005). Relearning sound localization with a new ear. *Journal of Neuroscience*, 25(22), 5413–5424. <https://doi.org/10.1523/JNEUROSCI.0850-05.2005>
- Van Wanrooij, M. M., & Van Opstal, A. J. (2007). Sound Localization Under Perturbed Binaural Hearing. *Journal of Neurophysiology*, 97(1), 715–726. <https://doi.org/10.1152/jn.00260.2006>
- Verdelet, G., Desoche, C., Volland, F., Farné, A., Coudert, A., Hermann, R., ... & Salemme, R. (2019, December). Assessing spatial and temporal reliability of the vive System as a tool for naturalistic behavioural research. In 2019 International Conference on 3D Immersion (IC3D) (pp. 1-8). IEEE.
- Visentin, C., Valzolgher, C., Pellegatti, M., Potente, P., Pavani, F., & Prodi, N. A comparison of simultaneously-obtained measures of listening effort: pupil dilation, verbal response time and self-rating. Accepted in *International Journal Of Audiology*.

- Vliegen, J., Van Grootel, T. J., & Van Opstal, A. J. (2004). Dynamic sound localization during rapid eye-head gaze shifts. *Journal of Neuroscience*, 24(42), 9291–9302. <https://doi.org/10.1523/JNEUROSCI.2671-04.2004>
- Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4), 339–368. <https://doi.org/10.1037/h0054629>
- Wanna, G. B., Noble, J. H., Carlson, M. L., Gifford, R. H., Dietrich, M. S., Haynes, D. S., ... Labadie, R. F. (2014). Impact of electrode design and surgical approach on scalar location and cochlear implant outcomes. *The Laryngoscope*, 124(S6), S1–S7. <https://doi.org/10.1002/LARY.24728>
- Warren, D. H. (1970). Intermodality interactions in spatial localization. *Cognitive Psychology*, 1(2), 114–133. [https://doi.org/10.1016/0010-0285\(70\)90008-3](https://doi.org/10.1016/0010-0285(70)90008-3)
- Wiebe, E. N., Lamb, A., Hardy, M., & Sharek, D. (2014). Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior*, 32, 123–132. <https://doi.org/10.1016/j.chb.2013.12.001>
- Wightman, F. L., & Kistler, D. J. (1999). Resolution of front–back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America*, 105(5), 2841–2853. <https://doi.org/10.1121/1.426899>
- Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (1991). Better speech recognition with cochlear implants. *Nature*, 352(6332), 236–238.
- World Health Organization. (2020). Deafness and hearing loss – Key fact. Retrieved from: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>
- Yost, W. A. (2017). Sound source localization identification accuracy: Envelope dependencies. *The Journal of the Acoustical Society of America*, 142(1), 173–185. <https://doi.org/10.1121/1.4990656>
- Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., & Tam, C. (2006). Perceptual recalibration in human sound localization: Learning to remediate front-back reversals. *The Journal of the Acoustical Society of America*, 120(1), 343–359. <https://doi.org/10.1121/1.2208429>
- Zeng, F. G. (2004). Trends in cochlear implants. *Trends in amplification*, 8(1), 1–34.
- Zimmer, U., & Macaluso, E. (2005). High binaural coherence determines successful sound localization and increased activity in posterior auditory areas. *Neuron*, 47(6), 893–905. <https://doi.org/10.1016/j.neuron.2005.07.019>
- Zonooz, B., Arani, E., Körding, K. P., Aalbers, P. A. T. R., Celikel, T., & Van Opstal, A. J. (2019). Spectral Weighting Underlies Perceived Sound Elevation. *Scientific Reports*, 9(1), 1–12. <https://doi.org/10.1038/s41598-018-37537-z>