UNIVERSITY OF GENOVA

PHD PROGRAM IN BIOENGINEERING AND ROBOTICS

Memory for sounds: novel technological solutions for the evaluation of mnestic skills

by

Walter Setti

Thesis submitted for the degree of *Doctor of Philosophy* (31° cycle)

December 2019

Monica Gori and Giulio Sandini	Supervisor
Luigi Cuturi	Co-Supervisor
Giorgio Cannata	Head of the PhD program

Thesis Jury:

Name, *University*	External examiner
Name, *University*	External examiner

Dibris

Department of Informatics, Bioengineering, Robotics and Systems Engineering

I would like to dedicate this thesis to my family which supported, advised and accompained me throughout these years

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Walter Setti
March 2020

Acknowledgements

And I would like to acknowledge ...

Abstract

Working memory (WM) plays a crucial role in helping individuals to perform everyday activities. The neural structures underlying this system continue to develop during infancy and reach maturity only late in development. Despite useful insights into visual memory mechanisms, audio-spatial memory has not been thoroughly investigated, especially in children and congenitally blind individuals.
The main scientific objective of this thesis was to increase knowledge of spatial WM and imagery abilities in the auditory modality. We focused on how these skills change during typical development and on the consequences of early visual deprivation. Our first hypothesis was that the changes in WM functionality and spatial skills occurring in the early years of life, influence the ability to remember and associate spatialized sounds or to explore and learn acoustic spatial layouts. Since vision plays a crucial role in spatial cognition (Thinus-Blanc and Gaunet, 1997), we expected blind individuals to encounter specific difficulties when asked to process and manipulate spatial information retained in memory, as already observed in the haptic modality (Cattaneo et al., 2008; Vecchi, 1998). Although some studies demonstrated the superior performance of the blind in various verbal-memory tasks (Amedi et al., 2003; Požár, 1982; Röder et al., 2001), very little is known on how they remember and manipulate acoustic spatial information.
The investigation of auditory cognition often requires specially adapted hardware and software solutions rarely available on the market. For example, in the case of studying cognitive functions that involve auditory spatial information, multiple acoustic spatial locations are required, such as numerous speakers or dedicated virtual acoustics. Thus, to the aim of this thesis, we took advantage of novel technological solutions developed explicitly for delivering non-visual spatialized stimuli. We worked on the software development of a vertical array of speakers (*ARENA2D*), an audio-tactile tablet (Audiobrush), and we designed a system based on an acoustic virtual reality (VR) simulation. These novel solutions were used to adapt validated clinical procedures (*Corsi-Block* test) and games (the card game *Memory*) to the auditory domain, to be also performed by visually impaired individuals. Thanks to the technologies developed in these years, we could investigate these topics and observed that

audio-spatial memory abilities are strongly affected by the developmental stage and the lack of visual experience.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

In recent years, interest has grown in the investigation of memory and mental imagery skills. The relationship between visuo-spatial cognition in general and other topics belonging to cognitive psychology like visual perception, working memory (WM), and executive/central processes is rather unclear (Denis et al., 2001). WM is a system devoted to the maintenance and sequential manipulation of incoming, task-relevant information involved in everyday tasks such as reasoning, learning, problem-solving, language comprehension, and motor planning (Baddeley, 1992; Baddeley et al., 1986). The brain structures underlying WM functionalities continue to develop during childhood till adolescence, and the maturity is only reached into young adulthood (Steenari et al., 2003). This system is involved in the generation of the mental images, "picture-like" representations that can be built up from multiple sensory inputs (e.g., tactile, visual, and auditory) (Dror and Kosslyn, 1994; Halpern and Zatorre, 1999). The role of vision in the construction of mental images is controversial. Recent studies on the haptic modality demonstrated that the generation of spatial representations is also possible in the absence of visual experience. However, congenitally blind people show limitations when asked to manipulate them actively (Vecchi et al., 2005; Wimmer et al., 2017). The studies in this context have been mostly carried out in the visual and haptic modalities. Due to the lack of proper technological solutions, audio-spatial memory abilities have not been deeply investigated yet. Nevertheless, audition is an important sensory modality, especially when vision is absent since birth. In such a clinical condition, hearing is fundamental to encode the distal environmental cues necessary to develop awareness, orientation, and mobility (Ungar et al., 1995).
In this introductory chapter, I will provide a brief summary of frameworks and definitions used to investigate the relationship between spatial cognition, mental imagery and WM. I will discuss the theoretical basis of these concepts and to which extent they are linked to

each other. The theoretical topics here discussed support the hypothesis that spatial and imagery processes are strictly linked to WM activity. Finally, I will give an overview of the technologies to date available and their limitations for the study of spatial skills in visually impaired individuals.

1.1 Spatial Cognition

Spatial cognition has a key role in everyday life because spatial skills are directly related to the individual's ability to move and navigate in the external world. When we move into the surrounding environment, we rely upon our motor functions to explore and understand the spatial properties of that environment. While moving, we automatically generate internal representations that are used to monitor and update the position of the body in space (Gilmore and Johnson, 1997; Wann et al., 1988). Regarding the relationship between motor functions and spatial representations, Milner and Goodale (Milner and Goodale, 2006) have assumed the existence of two different systems for the processing of visual and spatial information. A visuo-motor system that encodes and processes visual information for movement's planning and an alternative structure for the manipulation of the spatial information necessary to generate spatial representations. This hypothesis takes its roots in the pioneering study of Mishkin and colleagues (Mishkin, 1982). They proposed the distinction between a "what" system for the manipulation of visual information, and a "where" system for locating objects in space.

Spatial information can be coded either egocentrically or allocentrically (Foreman and Gillett, 1997). The first refers to a representation of space centered on the observers' body (e.g., the head or the hand). When instead the spatial representation is based on the mutual spatial relations among the items composing the surrounding environment regardless of their spatial relationship with the observer position, allocentric frames of reference come into play. Both frames of reference are closely linked to motor functions and spatial exploration skills such as grasping and reaching skills (egocentric) or in orienting and navigational abilities (allocentric).

Spatial cognition is especially involved in the ability to manipulate and maintain in memory spatialized items. In the following section, I will then describe the features of WM, focusing especially on the processing of spatial information and the role of development and visual experience.

1.2 Working Memory (WM)

WM refers to a framework of structures and processes involved in the temporary storage and manipulation of information retained in memory. This system is involved in complex cognitive skills, such as learning, reasoning, and motor planning. The majority of the cognitive tasks we accomplish every day (e.g., reading a newspaper article, comparing the different prices of a given product) usually require multiple steps with intermediate results. They must be kept in mind temporarily to perform successfully the task the individual is carrying out (Miyake, 1999). The WM construct was initially proposed by Baddeley (Baddeley et al., 1986; Baddeley and Hitch, 1974; Baddeley and Logie, 1999) and partly replaced the short-term memory construct. The WM, while maintaining the characteristics of the short-term memory (limited capacity and temporary maintenance of the information), represents a more complex system that allows the processing of information during the execution of cognitive tasks. Over the years, however, the concept of WM has undergone constant evolution, due to the numerous formulations and reformulations of the models (Baddeley, 2000; Baddeley et al., 1986; Cornoldi et al., 2000; Cornoldi and Vecchi, 2003; Engle et al., 1999; Miyake and Shah, 1999). Here, two of the most influencing and widely accepted models are presented:

1. Multi-Component model;
2. Continuum model;

1.2.1 Multi-Component model

This model was proposed for the first time by Baddeley and Hitch (Baddeley and Hitch, 1974). According to the original model (Baddeley et al., 1986), the WM system comprises a number of different subsystems, each processing different information (see Figure 1.1). The central executive component (CE) acts as an attentional control system and monitors two independent subsystems: the visuo-spatial sketchpad (VSSP) for the elaboration of spatial material and the phonological loop, mainly involved in the processing of verbal information (Baddeley, 1992; Baddeley et al., 1986). Two decades ago, a fourth component was added to the original model: the episodic buffer (Baddeley, 2000). This is a limited store of information that integrates phonological, visual and spatial material along with the information maintained in long-term memory. The phonological loop is further divided into a passive phonological store and an active articulatory-loop that refreshes the verbal material maintained in memory through a process known as rehearsal. Logie and Marchetti (Logie and Marchetti, 1991) suggested for

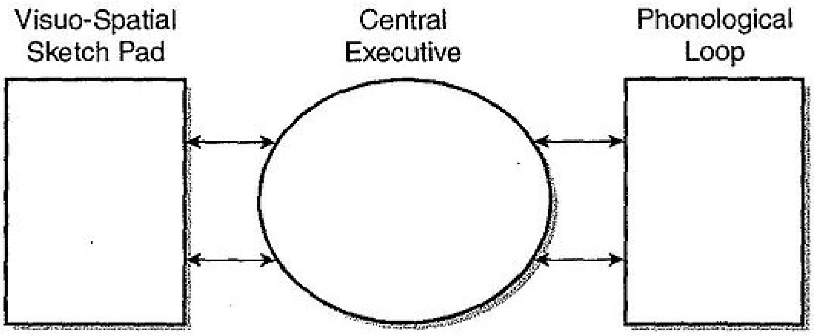


Figure 1.1 **First Model proposed by Baddeley in 1986**

the first time that the visual and spatial components of the VSSP are separated. In detail, the visual component (the *visual cache*) works as a passive storage of visual information. Conversely, the spatial component (the *inner scribe*) can be considered as a mechanism of active rehearsal that is the equivalent of the phonological-articulatory loop.

1.2.2 Continuum model

Baddeley (Baddeley et al., 1986) considers the subsystems composing the model as separate entities. However, the WM system and its components, rather than discrete entities, can be considered as distributed along a continuum (Conway et al., 2001; Cornoldi et al., 2000). Based on these evidences, Vecchi and Cornoldi developed a continuum model of WM (Cornoldi and Vecchi, 2004), which goes against the assumption of a single comprehensive system (Roncadin et al., 2007) or the multi-component model of Baddely (Baddeley et al., 1986) (see Figure 1.2). The authors proposed a cone-shaped model that develops along two dimensions: horizontal and vertical. The vertical size refers to the level of active control needed to accomplish a specific cognitive task, such as the level of attention. Distributed along the vertical continuum are the more automatic skills, in which the involvement of working memory is low, and the processes that involve the CE, which instead require higher active control. These processes are directly linked to the cognitive load needed to solve the task; the demand for cognitive sources indeed increases with the level of dynamic control required. The horizontal dimension is related to the type of material to be processed (e.g., verbal, visual, spatial), that is the nature of the perceptual input. Vecchi and co-workers argued that memory processes depend on the kind of information to be memorized and their horizontal distance on the continuum; the shorter the range, the stronger the interconnection between stimulus properties. At the basis of the cone-shaped model, the early processing of perceptual inputs is distinctly separated from each other. Although separated, visual and spatial material show contiguity going up the vertical dimension of the cone-shaped model, whereas linguistic objects maintain their distinction from the other components. This structure is in line with the division of Baddeley's model between two separate subsystems for the processing of phonological and visuo-spatial information (Baddeley et al., 1986). However, by focusing on visuo-spatial processing, both the multicomponent and the cone-

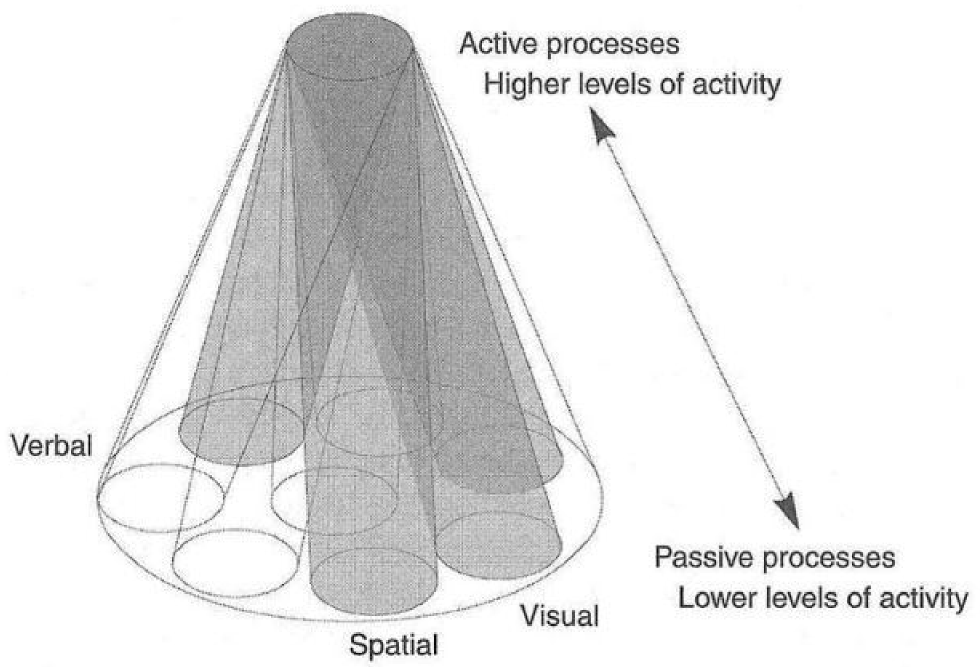


Figure 1.2 **Continuum WM model**

shaped model predict these components to be separated from each other at a lower processing level. At a cortical level, this aspect can be seen in the distinction between two neural pathways for the visual processing of object features and its spatial information: the ventral *what* paths and the dorsal *where* pathways (Goodale and Humphrey, 1998). The ventral pathway, from occipital to temporal cortex, is devoted to the processing of visual properties of the stimulus such as color, shape, and texture. The dorsal stream is activated by visuospatial information and goes from the occipital to the inferior parietal cortex (Mishkin, 1982). The same segregation between *what* and *where* is valid in the context of auditory perception, with some differences with respect to the visual modality. Firstly, in the visual system, information from one hemifield is sent to the contralateral hemisphere. On the contrary, in the auditory modality, both regions receive information from both ears (Romand and Varela-Nieto, 2014).

The visual cortex has a topographic organization, with adjacent neurons, responding to nearby stimuli in the retina (Kosslyn et al., 1993). Sound frequency is represented in a tonotopic map, and this representation is maintained across the whole auditory pathway, including the auditory cortex (Middlebrooks, 2002). Previous works on spatial WM pointed out that the dorsal visual pathway (prefrontal and post parietal cortices) is also involved in WM processing of both visual and auditory locations (Arnott et al., 2004; Curtis and D'Esposito, 2003; Martinkauppi et al., 2000). However, it is still an open debate about whether auditory and visual information are processed in separate brain networks or in a common neuronal network that integrates the two information.

1.2.3 The development of WM

During the course of life, the human brain changes. The primary motor and sensory areas are the first to mature. In contrast, the higher-order association areas (e.g., parietal and temporal areas), involved in advanced functions, reach maturity into adolescence and beyond it (Gogtay et al., 2006; Thompson et al., 2004). In parallel to neuroanatomical brain development, also behavioral and cognitive abilities develop during childhood into adolescence. Most of the WM functionalities indeed are ascribed to the prefrontal cortex that is one of the last brain areas to reach maturity. Infants (8 to 12 months old) can localize objects and remember their locations in the surrounding environment (Diamond, 1990). WM functions at that age are mostly ascribed to simple maintenance and retrieval of information. The WM structure is already in place by the age of 4 y.o. and the functionality of each component increases until early adolescence (Cowan and Alloway, 2009). The maturation of WM functionality can indeed be seen by the better performance in WM tasks with varying processing demands, from the age of 6 years-old into middle childhood (Byrnes, 2010; Gathercole et al., 2004). Several studies demonstrated that executive functions are at use, starting from the age of 8 years-old, but lack of functional integrity (Luciana, 2003). Until adolescence, indeed, even though older children exhibit superior performance in visual-spatial WM tasks compared to younger peers, the executive control on WM tasks is not as efficient as that of adults (Conklin et al., 2007); the first years of life mainly serve as tuning of perceptual and sensory-motor functions. Only later in development, complex processes associated with WM functions are integrated. A mature level of storage capacity is reached around the age of 11 y.o.; complex WM functions instead can be accomplished only after the age of 15 y.o. (Huizinga et al., 2006; Luna et al., 2004b). The age-related improvement in memory has to be also ascribed to the maturation of other cognitive abilities such as processing speed, attention, and storage

capacities (Cowan et al., 1992; Gathercole, 1999; Hulme and Tordoff, 1989).
The development of the phonological loop instead can be observed in the changes relative to the encoding of acoustic or semantic features of the words. Before the age of 8 years-old, the children recall the items in the surrounding environment by focusing on their physical features (e.g., shape, color, texture) (Henry, 2011; Kemps et al., 2000; Pickering, 2001). Thanks to the greater efficiency in the phonological loop, older children can complement visual information with phonological coding. Furthermore, before the 8th year of life, children tend to remember better words that are acoustically rather than semantically related to each other (Hasher and Clifton, 1974). In detail, the acoustic features of the words seem to be essential for the memorization process until the age of 8-9 years-old; afterward, the children better remember the words by considering what they refer to (e.g., the meaning or concept behind). Thus, the improvement in WM task performance has been associated with the development of the ability to re-code visually presented information into verbal form (Pickering, 2001). WM continues to develop from infancy into young adulthood, and the mature level of storage function is reached between 11-13 years of age while complex WM functions can be accomplished only after approximately the age of 15-19 years of age (Luna et al., 2004a).

1.2.4 Effect of blindness on WM abilities and mental imagery

One of the main functions ascribed to the WM system, especially the VSSP, is mental imagery (Cornoldi and Vecchi, 2003). This is a quasi-perceptual experience occurring in the absence of actual stimuli for the relevant perception (Finke and Freyd, 1989; Rinck and Denis, 2004). Mental imagery is directly involved in cognitive skills such as learning, memory (Yates, 1966), problem-solving, reasoning (Féry, 2003), and original and creative thoughts (LeBoutillier and Marks, 2003). The nature of these representations has been long discussed in the *imagery debate*. To date, the theory most accredited about the nature of the mental images is the one formulated by Kosslyn (Kosslyn, 1980). According to the author, the mental images are "picture-like" representations, as confirmed by paradigms such as mental rotation or mental scanning (Farah et al., 1988; Vingerhoets et al., 2002). In a mental rotation paradigm, the participants are asked to indicate whether two figures (where one is rotated relative to the other) are equal or different (Shepard and Metzler, 1971). Conversely, in a mental scanning task, the participants are asked to memorize spatial layouts and then to mentally scan the pathway from one location to another of the learned structure (Denis et al., 2001). Kosslyn's initial theory assumed that imagery processes partially overlap with

perceptual mechanisms.
However imagery should not be identified with visual perception. Visual perception (or representation for just perceived items) is mediated by bottom-up mechanisms that start in the retina. Conversely, imagery is influenced by the activation of spatial subsystems and higher-order visual areas. At least at the functional level, both perception and imagery share the same structure that is the so-called "visual buffer," which corresponds to the early visual areas. Imagery would result from a top-down activation of the cortical regions that had processed the information during perception (regardless of the input modality). Thus, imagery is based on a memory retrieval process by the recalling of the information stored in memory. Visual mental images can, therefore, be seen not as simple copies of perceptual inputs but as the end product of a series of constructive processes (Pietrini et al., 2004). Although most of the available studies have been carried out on visual imagery, partially overlapping networks for perception and imagery have also been found for tactile, olfactory, motor, and auditory imagery (Jeannerod et al., 1995; Kobayashi et al., 2004; Kosslyn et al., 2001). Visuo-spatial mental images can originate from a variety of perceptual inputs (e.g., visual, haptic and verbal). The results of the studies on congenitally blind individuals have been, therefore, fundamental in the context of the imagery debate. Visually impaired and blind individuals can generate and manipulate mental images, mostly thanks to long-term memory information, haptic exploration, or verbal description (Carreiras and Codina, 1992; Kerr, 1983; Lederman and Klatzky, 1990). For instance, blind individuals better remember concrete rather than abstract words (Cornoldi et al., 1979). Also, the representation of colors can be present in case of early visual deprivation (Marmor, 1978); color representation in the blind may indeed be associated with abstract knowledge. Some authors suggested that the images created by congenitally blind people are *visual*, mostly thank to the activation of the visual buffer by non-visual inputs (Maltini et al., 2003). Visual features, such as dimension, shape, or texture, can indeed be perceived through touch and reproduced in the internal images. Even though the "vividness" of the images generated by the blind is still an open debate, it is clear that blind individuals can represent spatial information. Representations of the blind are similar to those generated by the sighted, as shown by the similar patterns of performance exhibited in different visuo-spatial tasks (Carreiras and Codina, 1992; Kerr, 1983; Klatzky and Lederman, 1995).
However, even though the lack of early visual experience does not affect the ability to generate mental images, blind individuals meet specific difficulties and limitations when asked to accomplish visuo-spatial tasks (Beni and Cornoldi, 1988; Cornoldi et al., 2000, 2009; Cornoldi and Vecchi, 2003, 2004; Vecchi et al., 2005). The processing of visuo-spatial

information is slower in case of early visual deprivation. Although visually impaired people can represent sensory information in a pictorial format, they need more time to operate on these representations. For instance, the blind show great difficulties in using perspective in mental representations (Arditi and Dacorogna, 1988) or in elaborating the third dimension when it comes to learning a haptic spatial layout (Vecchi, 1998). Many limitations are also encountered by non-sighted people when the task request is to update the mental image continuously or to evaluate distances and inclinations (Juurmaa and Lehtinen-Railo, 1994). However, given that blindness does not influence the ability to generate mental images does this sensory deprivation impact on the actual manipulation of such representations? To answer this question, Cornoldi and colleagues (Cornoldi et al., 2009) asked blind and sighted participants to firstly generate and then to update the mental representation of a haptic pathway. The authors found a significant difference between the groups only when the request was to recall the whole pathway. They addressed this result to the difference between *passive* and *active* memory tasks. The first require only the maintenance of information in memory; the second instead require large sequences of mental manipulation. In other words, the authors concluded that, when the task only requires the maintenance in memory of spatial information, blind individuals perform as the sighted. Conversely, their performances drop when asked to manipulate this information actively.
To summarize, congenitally blind people can generate mental representations of objects or spatial layouts, and their spatial memory abilities are comparable to the ones of the sighted when the task demand is not high. However, due to the limitations imposed by early visual deprivation, they encounter difficulties when asked to process this information (e.g., in active memory tasks or when asked to manipulate several items at the same time). Only vision indeed allows the simultaneous processing of several items at the same time, while the simultaneous maintenance of several stimuli in memory is strongly affected by congenital blindness (Cornoldi and Vecchi, 2003).

1.3 Spatial Processes, Visual Mental Imagery, WM

As stated at the beginning of section 1, there are difficulties in using a common approach to study WM, spatial cognition, and mental imagery. The research on spatial cognition hypothesized a relationship between motor and spatial processes. Conversely, the research on imagery identified a link between perception and mental representations. According to Kosslyn (Kosslyn, 1980), the mental images are internal representations used in high-level perceptual processes such as planning and control of movements. An internal mental representation is needed to orient ourselves in the surrounding environment, coordinate motor activities, visualizing objects and scenes, or when the information about the external world has to be stored in a map-like format (Thinus-Blanc and Gaunet, 1997). Thus the generation of these mental images requires the integration of different sources of information. Imagery is identified as a WM function because the generation, the maintenance, and the processing of these representations are functions ascribed to the VSSP and CE components of the system. Analogously, representations used for motor and orienting activity are an outcome of the VSSP (Cornoldi and Vecchi, 2003). Therefore, the spatial representations used for spatial cognition or for moving in the surrounding environment, are generated in the WM system, by the involvement of perceptual inputs or from the information stored in long-term memory (see section 1.3). In the experiments presented in this thesis, I will describe to which extent the stage of development and the lack of early visual deprivation impact audio-spatial memory abilities and the processing of the mental representations generated after the spatial exploration of acoustic layouts.

1.4 Technological solutions for visually impaired people

Early visual deprivation affects psychomotor and cognitive abilities (Warren, 1994). Congenitally blindness strongly influences the typical development of important spatial and social skills (Cappagli et al., 2017b; Gori et al., 2013; Morrongiello et al., 1998), and usually, these impairments persist into adulthood. Thus, it is necessary to intervene in the first period of life to foster an increase in independence and quality of life in early visually impaired people.

1.4.1 Overview of the existing assistive devices

The design of novel technological solutions and their inclusion in clinical and eventually rehabilitative contexts is therefore essential to achieve these goals. The technologies developed in the last decades belong to the family of the substitution devices (SSDs). These apparatus convey visual signals through the other sensory modalities. However, there are very few products to date available on the market since many of them are still prototypes not usable by non-sighted individuals. Furthermore, the majority of these devices are very complex and, therefore, not usable by visually impaired individuals.

The SSDs nowadays available for blind individuals, mostly rely on the use of the sense of touch and audition. The latter is the most critical sense for orientation and mobility in the surrounding environment.

Conversely, touch is the best modality to explore non-audible items and to read through tactile maps (e.g., the Braille). SSDs devices usually replace the functions of a missing sensory modality with another one. For instance, they convert stimuli conveyed through a specific sensory modality (e.g., a visual stimulus) into a signal or another stimulus that can be accessible by another sensory modality (e.g., a sound or a haptic matrix). The majority of the SSDs to date available on the market include a sensor that permits the switch of a particular type of energy (e.g., light or sound) into an electrical signal. An example of an SSD is the cane. When the rod hits an object, the hand receptors are stimulated and give a clue on the location of the object (Bach-y Rita and Kercel, 2003). Assistive technologies that serve for orientation and mobility are mainly electronic travel aids (ETAs). They help blind people to create a representation of the surrounding environment by hearing or touch.

Navigation on the web for non-sighted people instead is made possible mostly with voice synthesizers that read on the computer or braille keyboards. In SSDs based on visual-to-tactile conversion, the optical image is transduced by vibratory or direct electrical stimulation. Bach-y-Rita developed the very first substitution device based on tactile stimulation in the 1960s (Bach-y Rita et al., 1960). This device converts a visual image recorded by a camera into a tactile image with a vibrotactile device, which is a matrix of 40 sensors worn by the user on the back of the chest (Collins and Bach-y Rita, 1973). Bach-y-Rita adopted this device to be placed on the tongue to transmit the visual information to the brain (The Tongue Display Unit (TDU)). The tongue is indeed very sensitive to touch.

Since Bach-y-Rita developed the first version of the visual-to-tactile sensory substitution device, many vibrotactile systems have been designed, ever smaller, and more portable (Lay-Ekuakille and Mukhopadhyay, 2010). After intensive training with these systems, visually impaired people improved in localization tasks (Jansson, 1983) and reported perceiving images in external space, instead of on the skin, suggesting that a process of externalization occurs with the extensive use of these substitution systems. The tactile devices are usually placed on regions of the skin that are rarely used and do not interfere with other sensory modalities. In everyday life, these tactile SSDs, thanks to the recent technological developments, are small energy-saving products. Nevertheless, these systems are costly. They need a sensitive region of the skin with a high density of mechanosensitive peripheral nerves to display complex information (i.e., the tongue). Nevertheless, the use of the sensor matrix on the tongue might cause irritation or even pain. Furthermore, this system is not comfortable and prevents the user from being able to speak.

Due to these limitations, many visual-to-auditory devices have been developed in these years. In this case, the mounted camera converts the visual frame into sounds sent through headphones. Most of these technological solutions rely on the use of echolocation or image-to-sound conversion (Auvray and Myin, 2009). Devices based on echolocation calculate the time taken by an ultrasonic signal to reach an object and return. These systems can give accurate information about the distance and directions of objects in the surrounding environment and might also be helpful for locomotion. One of the most famous devices belonging to this group is the SonicEye (Sohl-Dickstein et al., 2015). This system amplifies the echoes produced by ultrasonic sources to help the identification of objects in space. Another device that belongs to this group is the SonicGuide (Kay, 2000). This system translates ultrasound echoes onto audible sounds in both ears for navigating and scanning objects. In the case of image-to-sound conversion, the images recorded by the camera are transmitted to users via headphones. The pixels of the recorded video image is converted into audio frequencies. The upper locations in these images are converted into high-pitch sounds; conversely, the areas on the bottom to low-pitch sounds. The luminosity of the visual objects instead is strictly related to sound amplitude: bright white pixels corresponding to sounds presented at maximal volume and a dark pixel corresponding to silence. The first visual-to-auditory substitution device was the vOICe system designed by Meijer (Meijer, 1992). The vOICe scans and processes the image frames of the video recorded by the camera. This snapshot is scanned from left-to-right and converted into sound, played through headphones from left-to-right. The images are scanned at intervals of one second. The auditory equivalent is dynamic and spread out over time. After the vOICe, other solutions have been developed: the PSVA (Prosthesis for Substitution of Vision by Audition) developed by Capelle and his colleagues (Capelle et al., 1998), the SmartSight developed by CronlyDillon and his colleagues (Cronly-Dillon et al., 2000), and the Vibe developed by Hanneton and his colleagues (Auvray and Myin, 2009). The Vibe uses inter-aural disparity, and the PSVA uses a pixel-frequency relationship mapping. These visual-to-auditory devices are useful for object localization and help the accomplishment of spatial tasks (Auvray et al., 2007; Cronly-Dillon et al., 2000; Renier et al., 2005). Hearing is the right candidate as substituting sensory modality because the human auditory system can easily manage rapid changes in sound patterns.

Furthermore, the generation of auditory stimuli requires a small amount of energy, and the wearable devices only need simple instrumentation like headphones, webcam, and computers (Capelle et al., 1998). The use of the SDDs has been widely investigated. Their benefits have been proven at behavioral and cortical levels.

1.4.2 Limitations of the solutions to date available

However, the SSDs present several limitations. Firstly, none of the devices shown above can convey the exact information usually carried by the visual system. It is indeed challenging to translate a high-resolution image into an equivalent audio or tactile signal. For instance, transmitting an image via the audio modality raises the problem of transforming rich spatial information into a time signal, which requires either some form of image scanning or complex code. The SSDs stress one sensory modality (e.g., hearing or touch), and after a long sensory stimulation, the user might not perceive the signals coming from the device due to habituation mechanisms. As already stated above, these devices require intense training sessions that can be very difficult to be accomplished for users both because of potentially excessive physical and cognitive load. The majority of the SSDs are prototypes and cannot be used in everyday life yet. Very few devices have been tested, and the ones employed in experimental settings involved a small sample size that usually do not include visually impaired individuals. Furthermore, these technologies are only thought for adults and cannot be easily used by children. The main reason why assistive devices created for adults are not feasible for children is that they are based on artificial operating principles that young users might find challenging to learn. Finally, the principal limitation of SSDs is that they aim at substituting the visual modality instead of considering the nature of the input sensory modality and the influence of sensory deprivation. In sum, most of the SSDs lack of adaptability to visually impaired people; technological development often tends to neglect the impact of sensory deprivation on perceptual and cognitive abilities fundamental to readout the information conveyed by the SSD.

1.5 Aims of the thesis

Spatial WM and imagery processes have been studied intensively and for a long time, in the visual and haptic domains during the lifespan and in case of early visual deprivation. Although blind people need to rely on the auditory modality to process surrounding spatial information, not much research has focused on the use of sounds to assess spatial WM in blindness. During my PhD, I studied audio-spatial memory abilities in sighted and blind individuals of all ages. The scientific aims of this thesis were:

1. To investigate audio-spatial memory abilities during the lifespan and the influence of WM development in the context of spatial audition;

2. To study the effect of early visual deprivation on audio-spatial memory skills and on the processing of spatial locations provided by real or simulated acoustic sources;

The investigation of these issues sheds light on the relationship between vision and audition in the context of cognitive processes and spatial cognition during typical development and in the case of congenital blindness. The devices and the experimental paradigms developed in the context of this thesis, provide an appropriate tool for the evaluation of cognitive and spatial skills in congenitally blind and sighted individuals of all ages.

Chapter 2

Technological solutions developed for the evaluation of audio-spatial memory abilities

High-level cognitive abilities, in blind individuals, have been mostly assessed in the last decades in the haptic domain or with verbal non-spatialized stimuli. The majority of the technologies to date available for visually impaired people, as discussed in chapter 1, are assistive and cannot be used in clinical settings for the evaluation of spatial or cognitive abilities. They indeed just convey visual signals in other sensory domains, and they are too complex to be used by children. The three devices on which I worked on during my PhD aimed at studying audio-spatial memory skills in sighted and blind individuals of all ages through spatialized sounds. I worked on the software development of two audio-tactile devices named *ARENA2D* and Audiobrush (see Figures 2.1 and 2.11 respectively, sections 2.1 and 2.3) that allow the spatial emission of sounds through real speakers. Furthermore, I designed a system based on an acoustic VR (see section 2.2). During these three years, these solutions served as tools for the evaluation of audio-spatial memory abilities, but the final aim will be to use them in clinical settings for rehabilitative purposes.

2.1 ARENA 2D

ARENA2D is a novel device that provides auditory feedback (Cappagli et al., 2017b; Setti et al., 2019). This novel technological solution allows the serial emission of spatialized sounds. The hardware was developed by the Electronic Design Laboratory (EDL) Unit of the Italian Institute of Technology (IIT) in Genoa. Two conference proceedings have been published on *ARENA2D* (https://ieeexplore.ieee.org/abstract/document/8802160 and https://ieeexplore.ieee.org/abstract/document/8802172). *ARENA2D* is a vertical array of 25

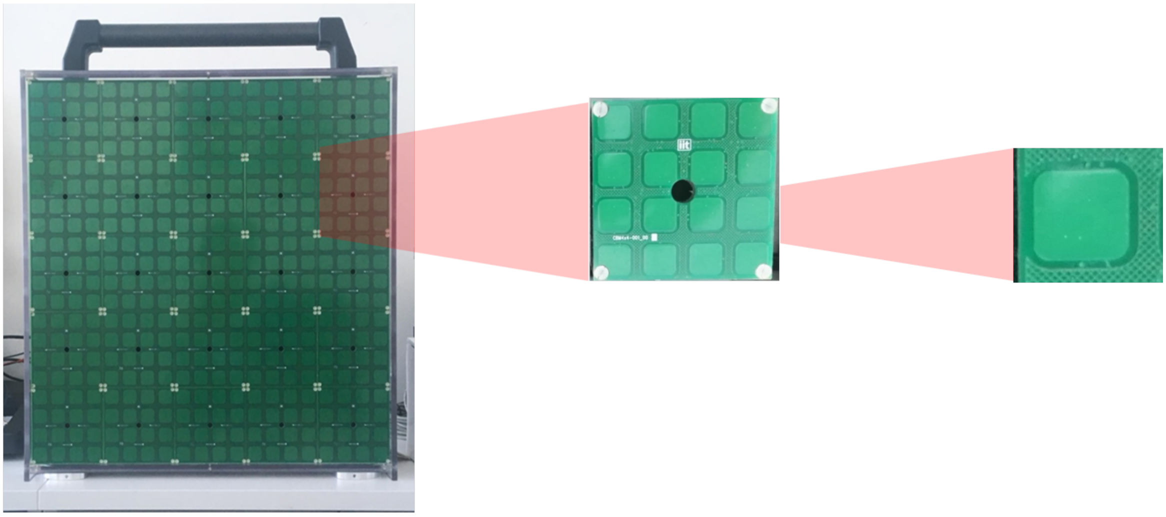


Figure 2.1 ***ARENA2D***. *ARENA2D* is shown at a different level of detail. The panel on the left presents the device. *ARENA2D* is a vertical surface (50 x 50 cm), composed of 25 haptic blocks, each provided of a loudspeaker in the center, arranged in the form of a matrix. The middle panel shows a detail of a haptic block. The black hole is the speaker from which the sound is emitted. The blocks are covered by 16 (4x4 matrix) tactile sensors (2 x 2 cm) that register the positions of the touches. The green square on the right represents a detail of a tactile sensor)

haptic blocks arranged in the form of a matrix (50 x 50 x 10 cm). Each block (5 x 5 cm) is a 4 x 4 matrix of tactile sensors or pads (2 x 2 cm each) provided of a speaker in its center. One of the most important hallmarks of this device is its modularity. The blocks are connected in cascade through USB cables and can be released from each other up to a distance of 20 cm. Furthermore, all the blocks can be programmed independently (for instance, the user can set different sounds for each block). *ARENA2D* weighs 12 Kg.

2.1.1 Block Diagram

In Figure 2.2 it is represented the general schema of the system: all the blocks composing *ARENA2D* are connected with wires through USB terminals. The connection with the PC is made possible by an interface, which is the *Host*, a serial network for data exchanging.

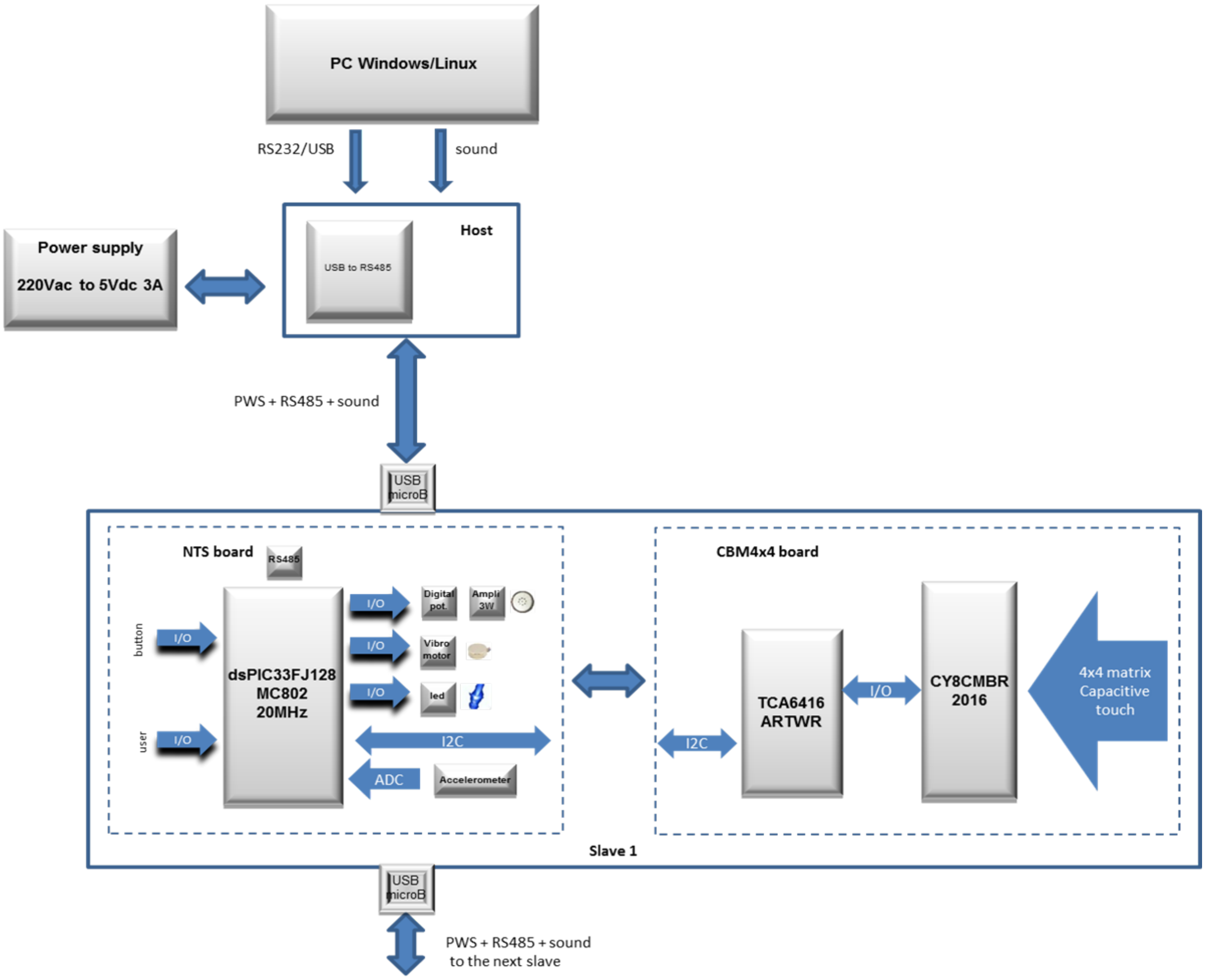


Figure 2.2 **Block Diagram**

2.1.2 Host Control

The *Host* units allow the PC and the *ARENA2D* to communicate with each other. The strings containing the commands are transferred by the PC via a virtual serial port and transferred to the chain of blocks through an RS485 interface. The interface ensures great reliability in terms of distance. Additionally, the USB port provides the necessary power for the unit.

2.1.3 Slave Unit

The blocks that compose *ARENA2D* consist of an NTS (New Tactile Sensor Unit) and CBM8x8 boards. The auditory signal, from the PC, reaches a class C amplifier (3 W). A potentiometer modifies the volume of the signal. The latter and the amplifier are controlled by a DSPIC33FJ128MC802 microcontroller that draws up the strings received by the host. In detail, the microcontroller interprets the commands received by the *Host*.

2.1.4 Wiring and Power cables

The system uses an external power supply (see Figure 2.3). All the NTS boards are connected

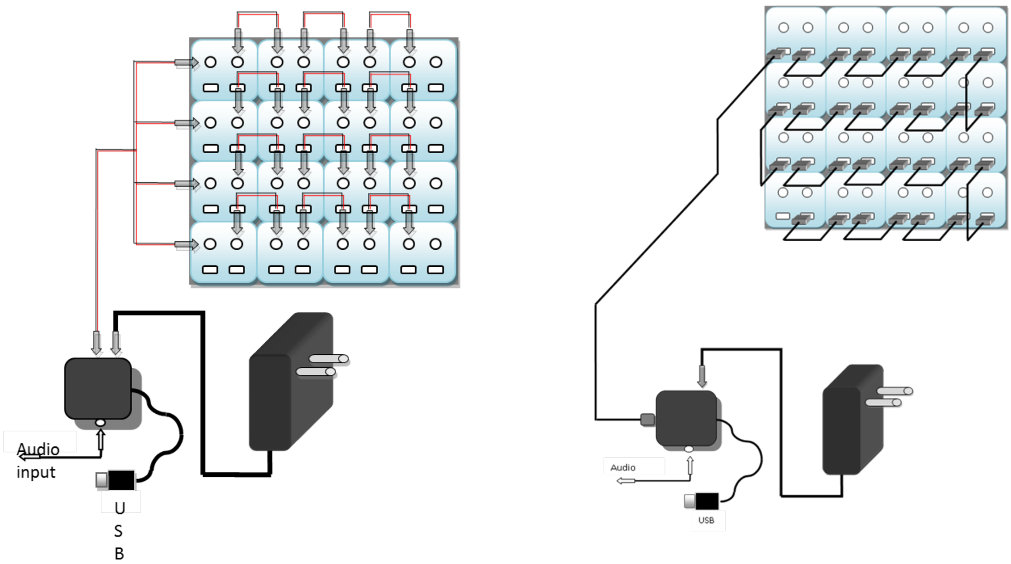


Figure 2.3 **Power and Data connections** The panel on the left shows the power connection and the panel on the right the data connections.

to the host through the data cable and also the power cord. Furthermore, the *Host* board is linked to the power supply via USB connection and an audio source (e.g., the PC).

2.1.5 Operating protocol

2.1.6 Serial Interface

The transport of information between the NTS boards and the PC is entrusted to a standard serial communication protocol (RS485). Data reception by the PC is made possible with an RS485/USB adapter. Driver's adapter uses a software emulation of a standard serial port (see Figure 2.4). In Windows, the serial port is found by using the windows "Device Manager," by looking at which serial port appears when the USB is connected to the PC.

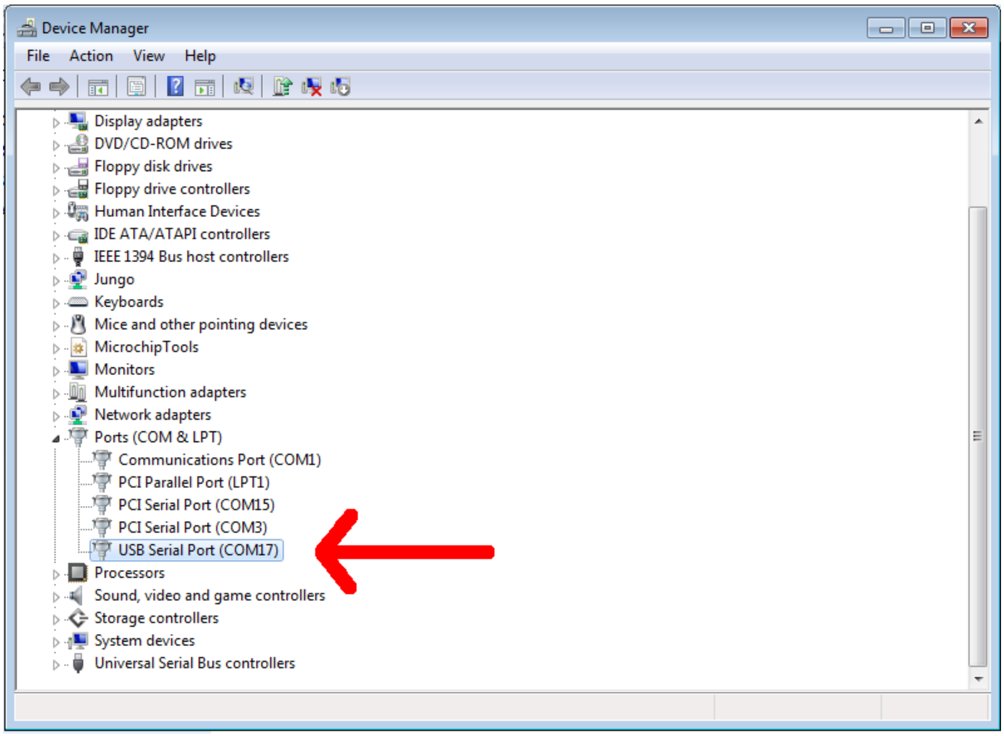


Figure 2.4 **Virtual serial communication.**

2.1.7 Tactile Sensors

The pads (see Figure 2.1, right panel) are capacitive sensors and belong to the CY8CMBR2016 CapSense Express capacitive touch sensing controller. This device supports up to 16 capacitive touch buttons that can be organized in any format, like a matrix array. Finally, the component costs are saved by the wide operation range (1.71 V to 5 V).

2.1.8 Communication Protocol

The communication is made possible by exchanging strings formatted as follows:

1. A starting byte (*Start*);
2. A byte that contains the NTS address to which the command is intended;
3. A byte that identifies the command;
4. Optionally, one or more bytes to identify the commands of the parameter;
5. An end byte;

The NTS boards complete the command after the receipt of the end byte. When one string has to be sent to all the boards, it is possible to use the broadcast address (0xFF recognized by all the boards). In this way, all the boards execute the string almost simultaneously. If the instructions to be sent are different, or the user does not want to command all the boards at the same time, a different communication protocol has to be employed. Several commands are therefore chained consecutively, and the ending byte will be sent just after the last command.

2.2 The audio virtual reality system

I spent three months at the Dyson School of Design Engineering at the Imperial College in London. There I designed a system based on an acoustic virtual reality (VR). This apparatus was developed to study audio-spatial memory abilities, using virtual auditory sources spatially displaced. Concerning *ARENA2D*, this system is cheaper and more portable and allows the displacement of sounds around the participant's head. The system consists of a software and a hardware part (an Arduino based keyboard), both described in the following paragraphs.

2.2.1 3D-Tune In application

The 3D Tune-In Toolkit was developed by the Imperial College of London and the University of Malaga (Cuevas-Rodríguez et al., 2018, 2019; Levtov et al., 2016). This project aimed at the design of a Toolkit to develop games and applications for hearing loss and hearing aid technology. The acoustic VR simulation is based on binaural spatialization, a technique that allows the creation of three-dimensional soundscapes using a simple pair of headphones as a playback device. The generation of binaural signals is based on the convolution of a monaural signal with head-related transfer functions (HRTFs), which model the directional filtering of the incoming signal due to the characteristics of the listener's body shape (i.e., torso, shoulders, head and pinna) (Algazi and Duda, 2010). In Figure 2.5, it is shown user's interface of the application. The Toolkit implements HRTF-based binaural spatialization integrating advanced features such as:

- User-imported HRTF. For this test, the HRTF number IRC1032 from the IRCAM Listen database was used (Caulkins et al., 2003; Kearney and Doyle, 2015);
- Barycentric interpolation between different Head-Related Impulse Responses (HRIRs), to allow smooth and realistic source/head movements;

- Customisation of the Interaural Time Differences (ITD) according to the head circumference of the listener;
- Advanced source distance simulation, including acoustic parallax and both near- and far-field correction filters;
- 3D reverberation based on a virtual-Ambisonic approach and Binaural Room Impulse Responses (BRIRs);

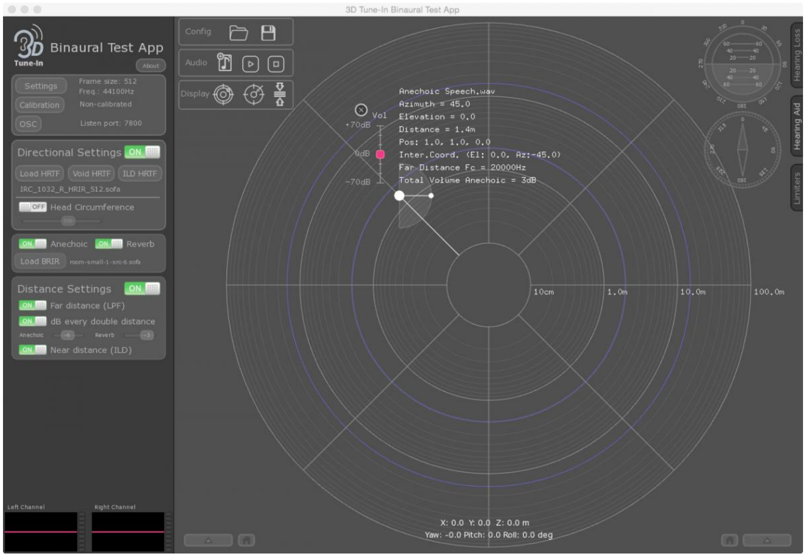


Figure 2.5 **Binaural Test App.** The picture, downloaded from the website of the project, represents the layout of the binaural test app. The app allows the spatialization of sound sources for binaural (headphones) listening. The app is also provided of some algorithms for simulating near-field and far-field auditory sources and reverberation in a 3D binaural context. Moreover, it allows the integration of hearing loss and hearing aid simulators. (http://www.3d-tune-in.eu/)

2.2.2 Communication protocol

Starting from the already developed 3D Tune-In Toolkit, we designed an application for blind individuals, by implementing the communication between Matlab (Matlab2017, The Mathworks) and the Toolkit. The 3D Tune-In Toolkit's test application receives instructions with OSC messages (Wright, 2005), based on wireless (WLAN) communication. The latter is a protocol for data exchange among devices based on the UDP communication protocol, which uses the IP addresses and the port of the units in communication. The UDP communication was established in Matlab by creating an UDP object (*udp()* function). Using the *oscsend()* Matlab function instead, it was possible to pilot the sounds when needed (e.g., start-stop individual sources, modify source positions according to head rotations, etc.), by using instructions compatible with the Toolkit.

2.2.3 Virtual Environment Design

Once the communication between Matlab and the Toolkit was established, we designed an acoustic virtual environment composed of virtual auditory sources. Our goal was to set a maximum number of sounds that could be easily perceived as coming from different locations. In other words, we wanted to deliver acoustic stimuli that could be easily distinguishable in terms of the location around the user's head.
To this aim, we carried out pilot studies where participants were presented with sound configurations and were asked to localize the positions of the sounds around their head. Sound configurations differed depending on the number of stimuli and the head (see Figure 2.6). The stimulus used was pink noise (2.5 seconds lasting, -25dB RMS). The pink noise was used because human sound sources localization abilities are improved when the signal that has to be localized is broadband (versus narrow-band - the wider the spectrum, the better) and when it has a complex envelope (e.g., it is not continuous) (Grantham, 1995; Moore, 2012). The Toolkit allows simulating simple attenuation for every double distance. The attenuation value is configurable (set to 6dB by default). On each trial, the sounds were presented one by one. After each emission, the participant was asked to indicate the position of the stimulus on the target configuration printed on paper. Each configuration was presented on a different block of trials. Twenty-five participants in total took part in the study. The percentage of correct responses for each sound stimulus was taken as a measurement of the ability to localize sounds in each configuration. As shown in Figure 2.6, the greater number of correct responses was achieved in the configuration shown in panel D of the figure (6 sounds, arranged onto 2 circles). In Table 2.1, the features of the single sources in this configuration (azimuth, elevation, distance from the listener) are shown. In the toolkit, each source is represented with two white dots (which turn to pink when audio is playing). The larger dot is connected to the listener with a white line (see Figure 2.5). The angle between the connection line and the horizontal axis of the chart indicates the azimuth angle, from 0.0 to 359.9 degrees anti-clockwise. The smaller dot represents the elevation of the source. The angle between the line that connects the two white dots and the horizontal axis of the chart represents the relative elevation of the source. The elevation varies between 0 and 90 degrees anti-clockwise (top-right quadrant), and clockwise from 360 to 270 (bottom-right quadrant).

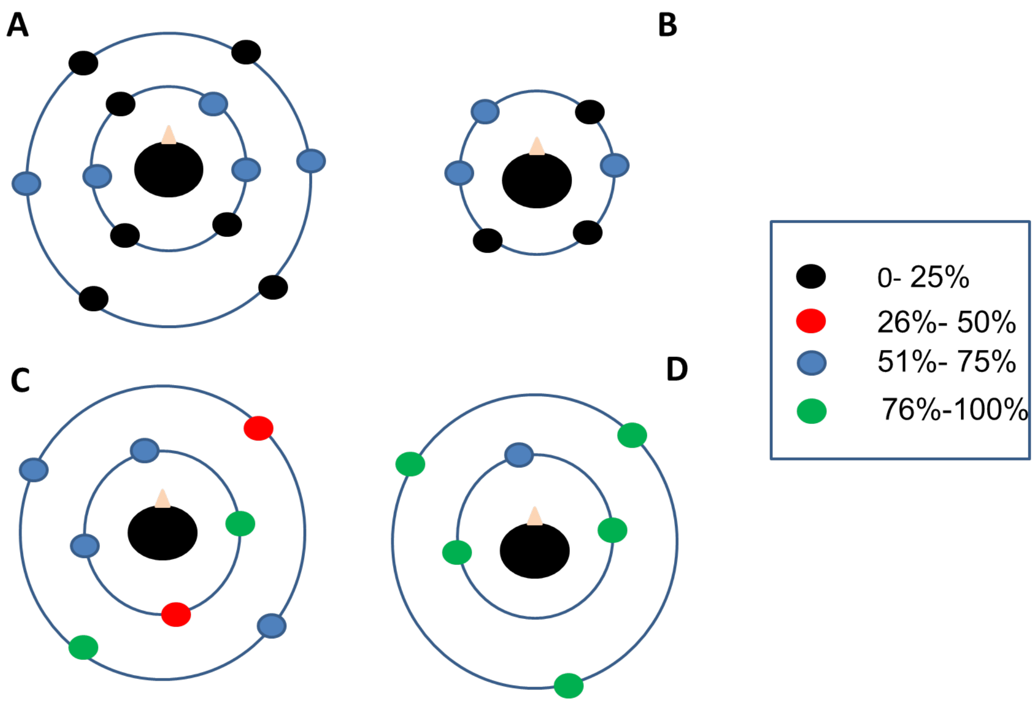


Figure 2.6 **Results of the pilot tests.** The filled circles represent the positions of the virtual sources. Participants listened to the sounds one by one, in random order, while looking at the configuration printed on a paper placed in front of them. In A C and D, the sounds were displaced on two circles: one closer to participants' head (perceived at a higher volume), the other further (perceived at a lower volume). Circle's colors represent the percentage of participants that correctly indicated the position of the sounds.

In the virtual environment, the sounds' elevation varied on a semi-circle. In order to represent the virtual environment more intuitively for the reader, in figure 2.7, azimuth, and elevation are represented relative to the participant's head. Thus the value in table 2.1 are related to this configuration. Some of the chosen sources have different elevations to improve localization of the sounds. Since The Head-Related transfer function (HRTF) is equal between the frontal and backward locations, we decided to position the sound in the southern location below the head to facilitate the discrimination between this sound source and the one in the north position.

Table 2.1 **Sources in the virtual environment.** In the table, the first column refers to the position of the sound for the head of the listener. The other three columns instead refer to the azimuth, the elevation of the sound and the distance of the sound from listeners' head in the virtual environment. The value of the elevation of the sound positioned behind user's head is equal to 327.5° in the virtual environment (according to the model used for the Toolkit). The value here presented corresponds to the model proposed in Figure 2.7

POSITION	AZIMUTH	ELEVATION	DISTANCE (m)
NORTH-WEST	59	25.4	1.2
NORTH	1.6	15.9	0.17
NORTH-EAST	318.6	15.9	1.3
EAST	261.2	0	0.19
WEST	104.9	0	0.18
SOUTH	191.1	237.5	1.9

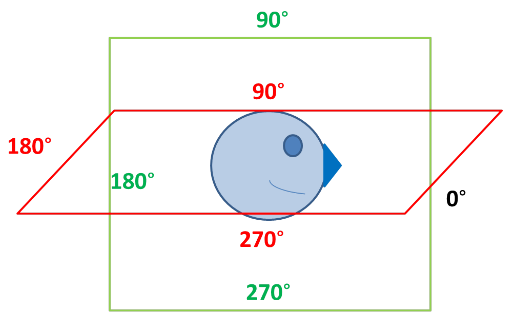


Figure 2.7 **Sound's azimuth and elevation.** The picture represents the radial variation of the azimuth (red square) and the elevation (green square) of the virtual sources. The 0° value is common to both azimuth and elevation and coincides with user's nose. The sound positions in the azimuth and elevation vary from 0° to 359.9° anti-clockwise.

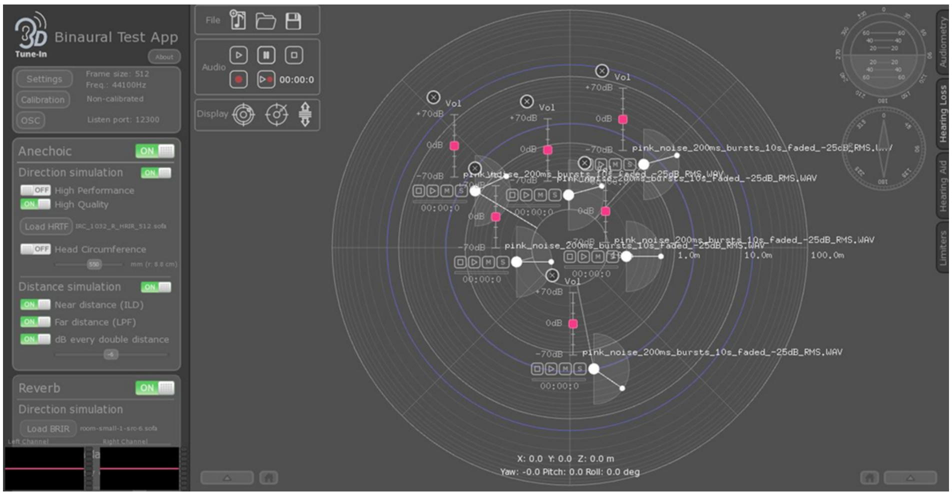


Figure 2.8 **Sounds' arrangement in the virtual environment**

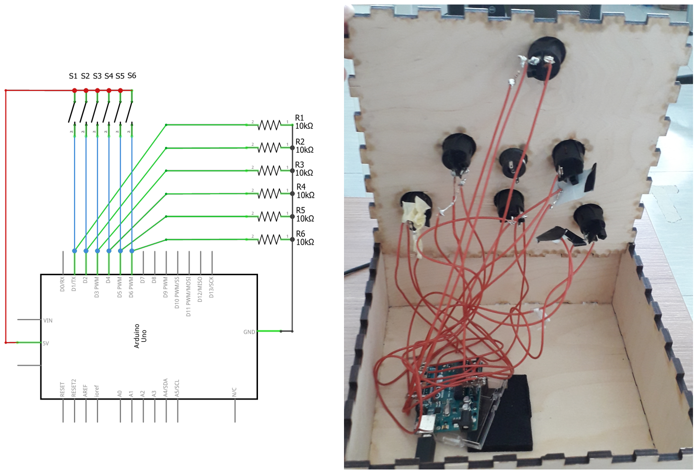
.

2.2.4 Arduino-Based keyboard

In order to register the responses in the experimental sessions, we designed an Arduino-based keyboard (see Figure 2.10). This system replicates the features of the virtual environment, to let the user know sounds' dispositions through haptic exploration. The keyboard, designed in AutoCad, was built as a laser-cut wooden box of 30x30 cm with a set of buttons on the top surface, which also hosted the needed electronic components (see Figures 2.9 and 2.10). The keyboard is composed of 6 buttons that replicate the positions of the virtual sources (see Figure 2.8) and a bigger central blue, which serves as a reference point.

The keyboard was controlled by an Arduino Uno[1], which logged the button presses by the user through a pull-down resistor configuration (see Figure 2.9a). The buttons were connected to the Arduino through tinned copper wires. The Arduino was connected to a computer via USB which allowed for serial communication and also powered the device. We used the Matlab support package for Arduino to allow communication between the readout and recording of participants' response. This package enables the user to acquire analog and digital data from the board over a USB connection or wirelessly over Wi-Fi. The *arduino()* function was used to connect the Arduino hardware to Matlab, by giving the port and board names as input data in the function. Once the *arduino* object is created, the user can program the hardware and employ custom add-on libraries.

[1]www.arduino.cc



(a) Keyboard circuit schematic (b) Keyboard circuit assembly

Figure 2.9 Keyboard circuit

(a) Top view (b) Side view

Figure 2.10 Assembled keyboard

2.3 Audiobrush

Audiobrush is a tablet (35 x 35 cm, see Figure 2.11) that allows serial and parallel emission of spatialized sounds.

The tablet is composed of 576 capacitive sensors. The auditory files have a fixed format: 8bit/sample, 22050 Hz, mono. The system is composed of six TAT1 and TAT2 boards (see Figure 2.12), managed locally by a Raspberry board (SO Linux) and through WiFi by a PC or a smartphone. The Raspberry looks for a WiFi network to connect with (named "AudioBrush"). The data are exchanged with the Raspberry at a fixed address. The TAT1 is composed of 96 tactile sensors (pads) and 12 speakers. It is connected to the TAT2

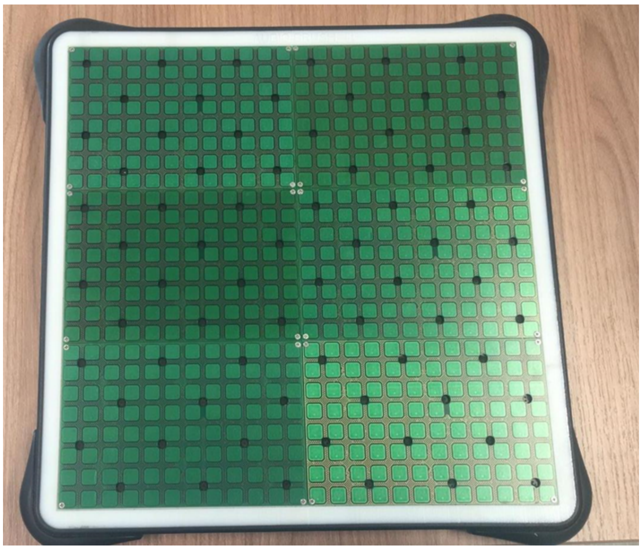


Figure 2.11 **Audiobrush**

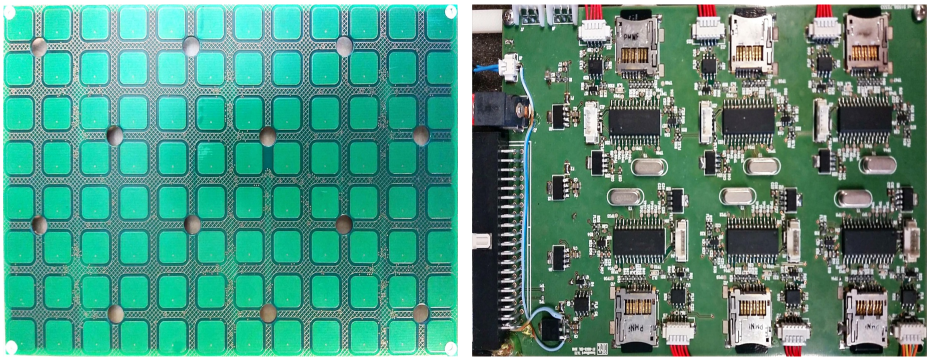


Figure 2.12 **TAT boards**. The panel on the left is the TAT1, while the panel on the right is the TAT2

that provided the power supply, the connection to RS485 port, and the auditory signal to be emitted. The TAT2 boards control the TAT1 boards but are also provided of voltage regulators for the whole device and RS232/RS485 converters.

2.3.1 Communication Protocol

The structure of the command strings is equal to the *ARENA2D* one. Every command transmitted by the Raspberry contains the address of the corresponding board. All the boards (TAT1, TAT2, and Raspberry) have their address. 0xFF is again the broadcast address. Before sending the reply, the boards wait for the end byte.
For both *ARENA2D* and AudioBrush, SSM2305 amplifiers were used. They are capable of

delivering 2 W of continuous output power (8 Ohms of impedance) and have a signal-to-noise ratio (SNR) > of 98 dB. The frequency range goes from 200 to 10 kHz, and the amplifier has a sensitive rating greater than 90 dB. The latency between the speakers depends on the code used to program the experiments. Conversely, the latency between the touch and the response is 10 ms in wired modality and 20 ms in wireless modality. Thanks to the virtual serial port, any dedicate software control (e.g., Matlab) or operating system can easily control the two devices.

2.3.2 Audio-Painting

Drawing is a visual art and is thought to be highly dependent on the visual modality (Likova, 2012; Thomas and Silk, 1990). Miller (Miller, 2015) stated that several visuo-spatial skills influence the development of drawing abilities: global representation of the spatial relations among the items in the environment, appreciation of lines and angles along with absolute and relative size. In the context of the current thesis, we focused on how defective graphics production can be correlated to a deficient development of visuo-spatial representations. This impairment, known as constructional apraxia (De Ajuriaguerra and Stambak, 1969), can be seen by looking at how children relate the individual part of a drawing into a unified whole. The observation about the accuracy of drawings and the order in which the single elements are drawn might constitute an important clue to shed light on the development of cognitive and spatial abilities (Taylor and Warren, 1984). Although it is typically assumed that drawing mostly relies upon vision, previous research indicates that congenitally blind people are also able to learn to draw (Kennedy, 1983; Kennedy and Juricevic, 2006; Ponchillia et al., 2008). Some congenitally blind individuals manifest advanced visual art skills, such as Esref Armagan. He is a Turkish artist that can draw and paint in colors even if he is an early visually deprived man (see Figure 2.13 to see two among his most famous paintings). However, even though some professional blind artists exist in these days, the recognition, reproduction, and creation of meaningful scenes are still extremely challenging for congenitally blind individuals. In the last decade, Likova and colleagues developed the so-called Cognitive-Kinesthetic Method (Likova, 2012; Likova and Cacciamani, 2018; Tyler and Likova, 2012), to use tactile drawing as a tool to investigate spatial and memory abilities in congenitally blind individuals. In this training, spread over five days, the participants learn how to draw complex and meaningful pictures by relying only upon the tactile domain. The persons involved in the training, firstly explore raised-line drawings with one hand. Afterward, they use the other hand to reproduce the picture from memory. Experimental


Figure 2.13 **Examples of Armagan's paintings**

results pointed out that drawing ability improves from before to after the training (Likova, 2012), highlighting enhanced spatial and memory abilities. Moreover, at the cortical level, the training leads to extensive cortical reorganization (Cacciamani and Likova, 2017; Likova et al., 2016). Particularly, these changes have also involved higher-level "memory" structures such as the hippocampus (Likova, 2015) and the perirhinal cortex (Cacciamani and Likova, 2016), as well as in brain connectivity (Cacciamani and Likova, 2017). Altogether, these studies indicate the effectiveness of drawing in congenitally blind individuals in the tactile domain. However, drawing, to our best knowledge, has never been used as a tool to investigate how blind individuals represent the external world, especially in the first period of life. In the training described above, the subjects had just to replace an already defined tactile painting. They were not allowed to draw a layout freely. Additionally, sounds have never been employed to draw, mostly for the lack of a technological solution capable of delivering spatialized sounds.

By means of Audiobrush (see description in section 2.3), we designed a novel paradigm to allow visually impaired children to represent the external world through drawing. Letting visually impaired children draw by using the auditory modality freely, leads to the development of innovative tools that can take advantage of audio-painting to evaluate and train memory abilities. The system might serve as a tool to test the awareness of the surrounding environment or of an observed/imagined scene and to teach the child how to build a proper representation of the external world. In the following paragraph, I am going to describe the proof of concepts of the application developed with Audiobrush.

2.3.3 Proof of concepts

To let the child draw with sounds, we used the keyboard of the auditory VR system (see section 2.2, figure 2.10) as the equivalent of a real palette and Audiobrush as the canvas for painting. The red buttons of the keyboard correspond to the color of a real palette. The procedure is illustrated in Figure 2.14. To date, only sighted children tried to draw with

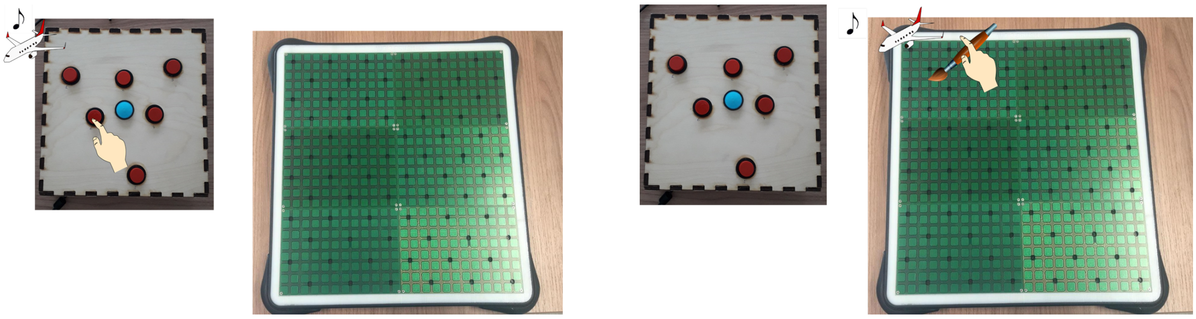


Figure 2.14 **Audio-Painting application.** The picture shows the application designed to teach how to draw to blind children. Initially, the child chooses the sound by pressing one of the buttons of the keyboard (panel on the left). Afterward, he takes a brush and, by swiping Audiobrush on the chosen position, the selected sound starts to be emitted (panel on the right)

Audiobrush. The procedure follows these points:

1. The children listen to the sounds by pressing the red buttons of the keyboard;
2. Once a sound is selected, they place a real brush on a specific position on Audiobrush;
3. The child swipes the brush on this position, and the selected sound starts to be emitted because the pads read the light touch of the brush;

The experimenter sees which board was selected by the child. Thus, from one of the speakers in that position, a sound will be emitted. The sounds composing the scene can be emitted one by one or simultaneously to have a correspondence between the drawing and the real external world. When all the sounds are reproduced simultaneously, the children have the impression of "looking through a window." Consequently, they can have a direct feedback of the plausibility of the drawing.

2.4 Experimental use of these solutions

The devices here presented have been used, during my three years of PhD, to design experimental protocols for the evaluation of audio-spatial WM abilities in sighted and blind individuals of all ages. In the following chapters, I will first describe the experiments designed with these technologies. We adapted the card game *Memory* to the auditory modality, to create a paradigm usable by children and adults with or without a visual disability. The system used to carry out the *Audio-Memory* was *ARENA2D*. This audio-tactile tablet was also employed to recreate meaningful spatial dispositions (see sections 3.1, 3.3 and 3.2) to evaluate the effect of the semantic information on memory performances in case of typical development or early visual deprivation. Finally, we adapted the *Corsi-Block* (Corsi, 1973) paradigm to the auditory modality to evaluate audio-spatial WM skills with the acoustic VR system.

Chapter 3

Investigation of audio-spatial memory abilities in children and blind individuals

The development of cognitive abilities (e.g., spatial WM memory) starts during the first years of life and continues throughout adulthood (Flavell, 1994; Rogoff, 1990). The identification of cognitive issues that might evolve in developmental delays has been the mission of pediatricians, psychologists, and educators alike (Chakrabarti and Fombonne, 2005; Robins et al., 2001). Traditional intellectual assessments comprehend cognition, attention, memory, and language, nowadays investigated and trained with games and online platforms (da Cunha et al., 2016; Folkvord et al., 2017; Giannakopoulos et al., 2018; Lee et al., 2018; Squire, 2003). Nevertheless, most of these solutions have a strong visual component, and they cannot be used to study these abilities in case of early visual deprivation.
To fill this gap, we took inspiration from the well-known card game *Memory*. The game works on attention, memory, concentration, and has many variations (da Cunha et al., 2016). In its original form, the *Memory* consists of a set of covered cards lying on a table. The goal of the game is to find the pairs among these cards. The participant turns the cards two at a time. If the two cards match, then the player takes them off from the table. Otherwise, they are set again on the covered side. The game ends when all the pairs are found. The *Memory*, and more in general the card games, can be played in two or alone and has many benefits (Beauchamp et al., 2016; Cao et al., 2016; Kuo et al., 2018; Raisamo et al., 2007). The principal ones are listed below:

1. To improve concentration;

2. To train visual and short term memory abilities;

3. To increase attention to detail;

4. To strengthen associations between concepts;

5. To help to classify objects that are grouped by similar traits;

We adapted this game to the auditory domain to study audio-spatial memory skills in adults and young individuals employing *ARENA2D* (see section 2.1). This device is indeed the most suitable for the design of this particular paradigm. It permits the spatial emission of several sounds (up to 25) that can be used as the corresponding *Memory* game cards. Furthermore, *ARENA2D* allowed us to design a game as close as possible to the real *Memory*, where the paired sounds can be removed from the game as the physical cards (see Figure 3.6 and 3.1 for the grids used in the *Audio-Memory*). The apertures of the grids correspond to the location of the sounds and can be covered by cardboard squares once the sounds are found and paired. Thanks to the features of *ARENA2D*, we have been also able to investigate how the participants explored the acoustic layout.

3.1 The role of development on audio-spatial memory abilities

The prefrontal cortex, directly involved in WM processes, is one of the latest cortical areas that reach maturity (Martinkauppi et al., 2000; Rypma and D'Esposito, 1999). Research on children highlighted that the tripartite WM model proposed by Baddeley (Baddeley and Hitch, 1974) is already in place at 4 years of age, and the capacity of the subsystems increases until the age of 12 (Alloway et al., 2006). After the 8th year of life, children start to use the subvocal rehearsal to memorize visual items (Henry, 2011; Miller, 2015). Regards audition, before the age of 8 years-old, children better recall acoustically related words in free-recalling tasks. With increasing age, words semantically related are better-memorized (Hasher and Clifton, 1974). Finally, the development also influences the ways of exploration and localization of items spatially displaced in the environment, at least in the auditory modality (Blades and Spencer, 1994). Most of the previous works on these topics have investigated WM skills or spatial exploration strategies through the visual channel or verbal, not spatialized sounds. Thus, very few results are available to shed light on these mechanisms in the acoustic domain. Even though previous research demonstrated that visuo-spatial and audio-spatial WM memory abilities improve with age (Vuontela et al., 2003), developmental changes in the ability to process complex auditory spatial arrangements and the influence of the semantic information have received little attention. In the current research, the scientific questions we wanted to address were:

1. To what extent does the development in WM functionalities (i.e., the phonological loop) affect audio-spatial memory abilities during the first period of life?

2. Is there a developmental trend in the exploration of acoustic spatial layouts?

We designed an experimental paradigm (named *Audio-Memory*) to investigate these questions in typical children aged from 6 to 11 years of age. The participants were asked to match sounds arranged on *ARENA2D* surface. The test was divided into two experimental conditions, named "call-call" and "call-name." In the first, the children paired two animal calls (e.g., two barks), while in the second, they had to match the call with the name of the corresponding animal (e.g., the bark with a recorded voice saying the name "dog"). This work provides an essential contribution to the study of memory abilities across age and unveils the role of acoustic and semantic binding in the context of auditory spatial processing. We hypothesized that the changes occurring in items encoding (i.e., the use of phonological codes), in words' clustering and spatial abilities during the first period of life, influence the capacity to memorize and pair the sounds displaced on *ARENA2D*. Thus the same changes and limitations occurring in the visual domain should also be valid in the auditory modality, especially in the youngest participants (e.g., 6-7 year-olds).

3.1.1 Material and Methods

Sample

Forty-four children took part in the experiment, divided into three groups based on their age: 6-7 year-olds (n = 18, 9 females, mean age: 6.39, std: 0.5), 8-9 year-olds (n = 12, 4 females; mean age: 8.54, std: 0.52), 10-11 year-olds (n = 14, 9 females, mean age: 10.2, std: 0.41). The participants did not have any cognitive or sensory impairment. The ethics committee of the local health service (ASL3, Genoa) approved the study, and all the experiments were conducted following the Declaration of Helsinki. Parental informed written consent was obtained, and all the children agreed to take part in the experiment. The study recruited the participants from the "Istituto Anna Frank" in Genoa. Before entering the experimental room, each child was informed he/s could interrupt the experiment at any moment. Nevertheless, no one asked for a pause. The choice of this age slots is because the 8th year of life has been considered as a critical age for the development of cognitive and spatial skills (Gathercole and Alloway, 2008; Gathercole et al., 2004), at least in the auditory domain.

Table 3.1 length of the spoken words used in the call-name condition.

ENGLISH WORD	ITALIAN WORD	DURATION (sec)
DOG	CANE	0.65
COW	MUCCA	0.8
DONKEY	ASINO	0.85
HORSE	CAVALLO	1
SHEEP	PECORA	0.7
ROOSTER	GALLO	0.9

Setup and Stimuli

The sounds for the experiment were emitted through *ARENA2D* (see section 2.1). We placed a cardboard grid on the surface of the device (Figure 3.1), leaving only 12 open squares measuring 4 x 4 cm with a small aperture (0.5 x 0.5 cm) in the center, to allow audibility of the central speaker. To better identify stimuli positions, we made apertures smaller than the single haptic block (Figure 3.1). We selected 6 target animal calls based on pilot studies that indicated these sounds to be easily understandable by the children (see table 3.1). *ARENA2D* was vertically oriented and placed right in front of the children. The animal calls were mp3 sounds downloaded from a royalty-free website (https://freesound.org/), each lasting 2.5 s. The sound pressure level (SPL) was maintained at 70 dB and the RMS (Root Mean Square) level was calibrated to be the same across the various signals. The time lengths of the spoken words were different from each other. The experiment was carried out in Italy, and the spoken words were in Italian (see Table 3.1). The children sat at a distance of 30 cm from the device; none of them had seen it before the experimental session. Besides, before entering the experimental room, they were blindfolded with blackened swim goggles positioned over a sleeping mask. The experimenter adjusted the subjects' position so that their eyes were aligned with the central haptic block of the device. The experimenter ensured the child was able to reach the corners of *ARENA2D*. Finally, during the whole experiment, the participants remained seated.

Experimental Procedure

The paradigm followed the rules of the card game *Memory*. It was divided into two experimental conditions, based on the associations the participants were asked to accomplish: call-call and call-name. In the call-call condition, we asked the child to pair animal calls (e.g., two bellows) while in the call-name the animal call with a recorded voice saying the animal's name (e.g., the bellow and the spoken word "cow"). The experimental conditions were presented in blocks and were counterbalanced across subjects (half of the children started with the call-call condition, the others with the call-name). Sounds' dispositions are shown in Figure 3.1. As the original *Memory*, the test ended when the child found all the

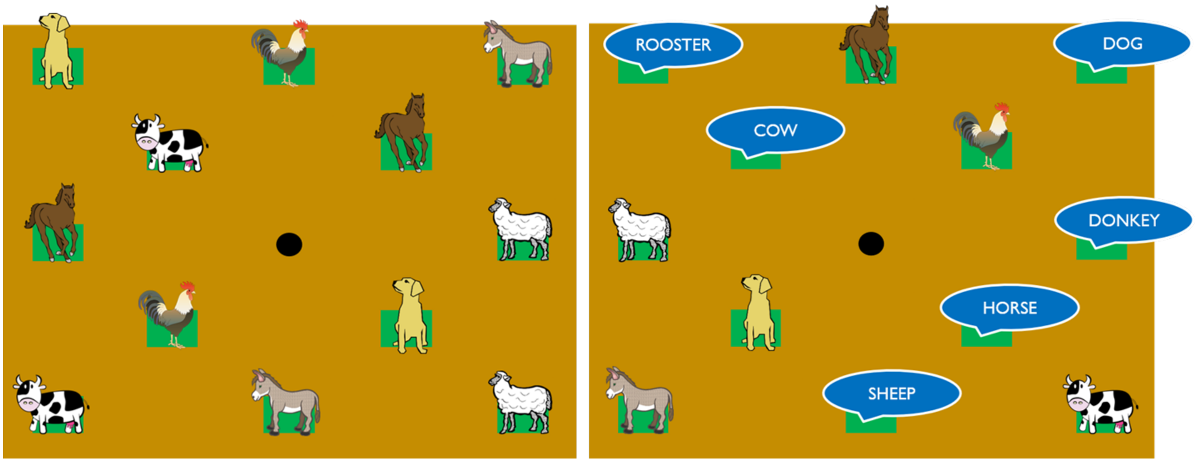


Figure 3.1 **Experimental Conditions**. The figure shows the stimuli used in both experimental conditions. The images on the figure, downloaded from a royalty-free images web archive (https://publicdomainvectors.org/), refers to the sounds emitted from the blocks once touched by the child. The panel on the left shows the sounds used in the call-call condition. The panel on the right instead, refers to the call-name condition.

pairs. To find the pairs, the participants used the index finger of the dominant hand to touch two haptic blocks per trial. After each touch, a sound was emitted from the selected speaker. The child was asked to hold in memory the location of the stimulus. Before touching another block, the participant had to wait for the sound to finish. In the classical *Memory*, once a pair is found, the cards are removed from the table. In our test, every time two stimuli were matched, the experimenter placed a cardboard cover on the apertures over the blocks. If the two sounds matched, a "TADA" sound was emitted from the central speaker as feedback sound. Otherwise, an error-like sound was emitted. However, even though the experimenter covered paired positions, the children could accidentally touch two already coupled locations. In these cases, a registered voice saying "NO" from the central block, alerted the child. We removed these attempts from the following analyses. Before the start of the test, the experimenter helped the child to familiarize themself with *ARENA2D*. The experimenter guided the child's dominant hand over the cardboard grid, and they counted the open slots; then, the child freely explored the device with both hands for 2 minutes. Afterward, the experimenter explained the rules of the test by recalling the rules of the *Memory*. Participants were instructed to pay attention to sounds' locations during the exploration of *ARENA2D*, trying to couple the pairs in the shortest time possible. The experimenter informed the subjects that when they touched a haptic block, a sound would have been played: either a call or a word. Subsequently, the children listened to the calls one by one. Before the beginning of each experimental condition, the experimenter conducted a practice session with the children to help them to understand the task. In these sessions, the experimenter guided the index finger of the dominant hand first on two unpaired and then on two paired speakers. The stimuli used in the practice sessions belong to the sounds used in the test session but, in the main experiment, they occupied different positions. Furthermore, the children listened to the feedback sounds for correct or incorrect stimuli. The test did not have a fixed duration because it was self-paced; however, on average, the test session lasted approximately 20 minutes for each child.

Data Analyses

To quantify children performances, we defined three parameters, related both to memory and exploration aspects:

1. *Score*: The score is a parameter that considers the number of touches on the same speakers (see Figure 3.2). The more the children touch the same block, the less the score they achieve on total trials. The score is an index that quantifies the overall performance in the test. The score was calculated as follows:

    - If both the blocks in an attempt are touched for the first time, the total score is not increased either decreased. This because memory processes do not influence these first touches;
    - If in an attempt, just one of the two blocks has already been touched, the total score is decreased by 1;
    - If in an attempt, both blocks have already been touched, the total score is decreased by 2;
    - When a pair is found, regardless of the number of touches per each block, the total score is increased by 10;

    When the speakers touched in an attempt have already been touched, the score was decreased. The participants indeed failed to remember the locations of the sounds, despite having already discovered them. See Figure 3.2 for details.

2. *Attempts to pair the stimuli*: With this parameter, we quantified how many attempts the children employ to pair two twin sounds once their positions have been discovered on the device. The higher the value, the more attempts are needed to couple the sounds, and therefore, the worse the performances. This parameter is fundamental to quantify the ability to maintaining the spatial locations of the sounds in memory.

3. *audio-anchor*: The audio-anchor serves for the analysis of the exploration strategy. The audio-anchor is an index that accounts for how many consecutive attempts the child makes by touching the same haptic blocks (see Figure 3.3 for details). The higher the value, the more the child adopts this exploration strategy. Suppose that the child, while exploring *ARENA2D*, finds the bellow and the bray. If the first block touched in the subsequent attempt is the bellow, the audio-anchor index increases by one. The index continues to increase until the child starts an attempt by touching another block. This parameter indicates which strategy the children use to build their spatial representation of sounds' locations.

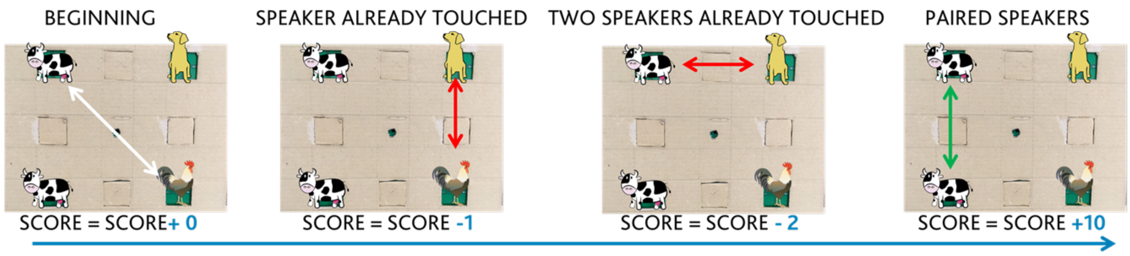


Figure 3.2 **Score, example of calculation**. The score is an index that decreases when the children return on the speakers previously touched. When they touch two blocks for the first time, the score equals 0. If they have already touched one or both speakers, the score decreases by one or two, respectively. When a pair is found, the score increases by ten. In the example, if the starting value were equal to zero, the final score would have been: 0 - 1 - 2 + 10 = 7. Depicted animals were downloaded from a royalty-free images web archive (https://publicdomainvectors.org/).

3.1.2 Statistical analyses

The three parameters listed in the Data Analysis section were analyzed separately. We first checked if the data followed a normal distribution or not with the Jarque-Bera test. We then ran a 2-ways (3x2) mixed measures ANOVAs, with *Group* (either 6-7, 8-9 and 10-11 year-olds) and *Condition* (either call-call or call-name) as between and within factors

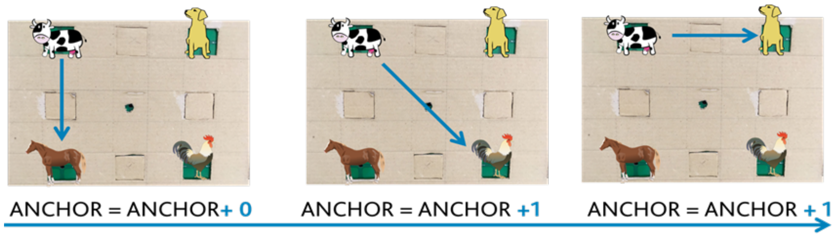


Figure 3.3 **Example of audio-anchor calculation.** The index equals zero at the beginning of the test. The index increase correlated with more attempts started with the same speaker. In the represented example, the final value would have been: 0+1+1=2. Depicted animals were downloaded from a royalty-free images web archive (https://publicdomainvectors.org/)

respectively. Two-tailed Student's t-tests (paired and unpaired) were used in the post-hoc analyses. *Bonferroni* correction was used to correct for multiple comparisons ($p<0.05$ was considered significant). The effect sizes are also reported (Cohen's d).

3.1.3 Main results

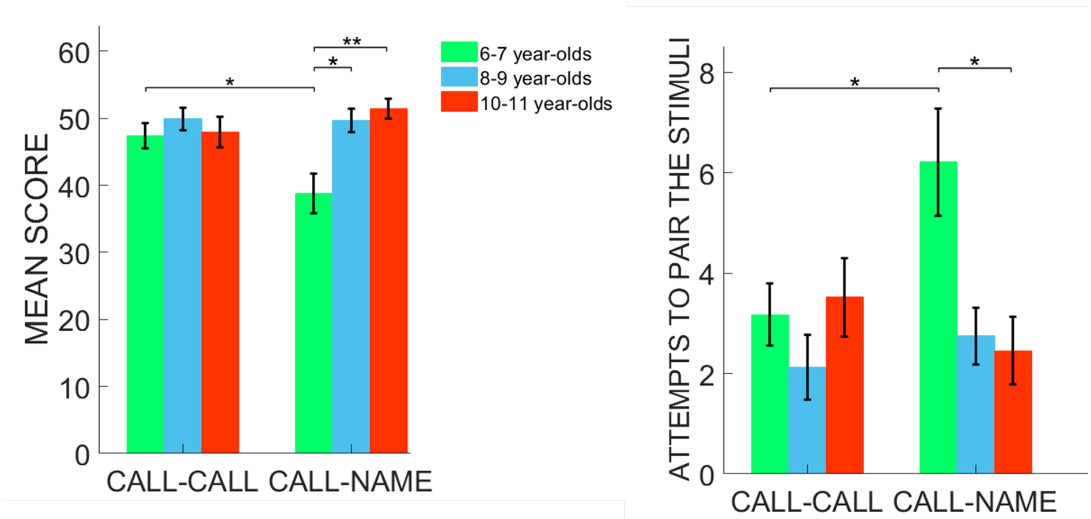


Figure 3.4 **Performance results.** The panel on the left refers to the score, while the panel on the right to the number of attempts. In the call-name condition, 6-7 year-olds reach a lower score in a higher number of attempts compared to the other two groups and to the call-call condition. One asterisk (*) represents $p<0.5$ Two asterisks (**) represent $p<0.01$ and three asterisks (***) $p<0.001$.

Analysis of the score revealed that 6-7 year-olds reach a lower score in the call-name condition, compared to the other two groups and the call-call condition (see Figure 3.4 panel on the left). The results show a significant main effect of the *Group* ($F(2,41)=5.51$,

$p<0.01$, Cohen's $d=0.62$) no significant effect given by the *Condition* ($F(2,41)=1.22$, $p=0.28$, Cohen's $d=0.17$), and a significant interaction *Group***Condition* ($F(2,41)=5.26$, $p<0.01$, Cohen's $d=0.44$). The follow-up ANOVAs (one-way ANOVA with *Condition* as within factor), show no significant difference between the groups in the call-call ($F(2,41)=0.39$ $p=0.67$, Cohen's $d=0.14$) but in the call-name ($F(2,41)=8.81$ $p<0.001$, Cohen's $d=0.53$) condition. As confirmed by the results of the post-hoc analyses, youngest participants reached a lower score in the call-name condition compared to the 8-9 (unpaired Student's t-tests: $t=2.78$, $p=0.03$, $df=28$, Cohen's $d=0.94$ *Bonferroni* corrected) and 10-11 (unpaired Student's t-tests: $t=3.48$, $p<0.01$, $df=30$, Cohen's $d=0.81$ *Bonferroni* corrected) year-olds and to the call-call condition (two-tailed paired Student's t-tests: $t=2.76$, $p=0.013$, $df=17$, Cohen's $d=0.54$ *Bonferroni* corrected). Regarding the number of attempts, the results show that, once a child discovers both the call and the corresponding name on *ARENA2D*, 6-7 year-olds need more attempts to pair them (see Figure 3.4, panel on the right) when compared to the oldest peers. We observed a significant main effect of the *Group* ($F(2,41)=4.5$ $p=0.017$, Cohen's $d=0.48$), no significant effect of the *Condition* ($F(2,41)=0.8$, $p=0.57$, Cohen's $d=0.2$) and a significant interaction between the two factors ($F(2,41)=3.82$, $p=0.03$, Cohen's $d=0.44$). Follow-up ANOVAs (one-way ANOVA with *Condition* as within factor) showed a significant interaction in the number of attempts to pair two discovered stimuli not in the call-call ($F(2,41)=0.98$, $p=0.39$, Cohen's $d=0.12$) but in the call-name condition ($F(2,41)=6.08$, $p<0.01$, Cohen's $d=0.48$). As shown in Figure 3.4, 6-7 year-olds need more attempts compared to the oldest participants (unpaired Student's t-tests: $t=2.79$, $p=0.023$, $df=30$, Cohen's $d= 0.89$, *Bonferroni* corrected). Furthermore, their performance were greater in the call-call experimental condition (paired t-test, $t=2.79$, $p=0.027$,$df=17$, Cohen's $d=0.57$ *Bonferroni* corrected). Concerning the use of the audio-anchor (see Figure 3.5), we found that regardless of the tested condition, 6-7 year-olds returned more to the same speakers when starting consecutive attempts than the other two groups. The analysis of variance highlights a significant main effect of the *Group* ($F(2,41)=10.94$, $p<0.001$, Cohen's $d=0.75$) and no significant effect of *Condition* ($F(2,41)=0.87$, $p=0.51$, Cohen's $d=0.27$) nor interaction between the two factors ($F(2,41)=0.6$, $p=0.35$, Cohen's $d=0.2$). Post-Hoc analyses revealed that, regardless of the experimental condition, 6-7 year-olds rely more on the audio-anchor strategy to explore the device compared to 8-9 (unpaired Student's t-test $t=3.09$, $p<0.01$, $df=58$, Cohen's $d=0.73$ *Bonferroni* corrected) and 10-11 (unpaired Student's t-test $t=2.72$, $p<0.001$, $df=58$, Cohen's $d=0.95$ *Bonferroni* corrected) year-olds.

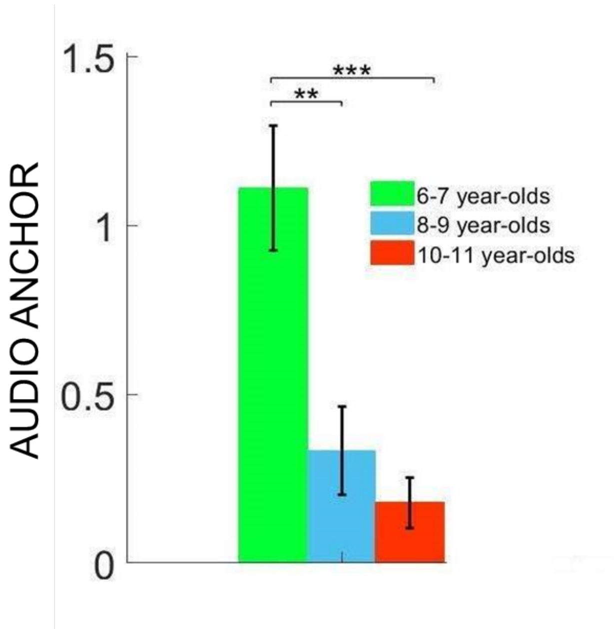


Figure 3.5 **audio-anchor** Data are presented as the mean across participants per each group; error bars represent the standard error. Regardless of the condition, 6-7 year-olds used this exploration strategy more than the other groups. Significant comparisons between groups are represented. Two asterisks (**) represent p<0.01 and three asterisks (***) p<0.001.

3.1.4 Discussion

In this work, we compared audio-spatial memory skills and spatial exploration strategies in children aged from 6 to 11 years-old, divided into three groups: 6-7, 8-9, 10-11 year-olds. They were required to pair two identical animal calls (e.g., two bellows) or the call with a registered voice reproducing the name of the animal (e.g., the bellow and the spoken word "cow"). Results show that the youngest participants need more attempts to pair the sounds and reach a lower score in the call-name condition compared to older peers and the other condition (see Figures 3.4). Moreover, they tend to start consecutive attempts by touching the same speakers and rely more on the strategy of the audio-anchor (see Figure 3.5).
Previous research demonstrated that auditory and visual WM performance increase with age (Gathercole, 1999; Kemps et al., 2000; Luciana and Nelson, 1998). WM skills indeed continue to develop until young adulthood, and the adult level in WM performance is only reached after 15 years of age (Huizinga et al., 2006; Luciana et al., 2005). The results of this study support the hypothesis that the 8th year of life is crucial for the encoding and memorization of visual objects (Casey et al., 2000; Klingberg, 2006; Kwon et al., 2002). After the age of 8 years-old, children start to rely on memorization strategies that make use of the sub-vocal rehearsal when it comes to immediate visual memory and verbal tasks (Henry,

2011; Miller, 2015). The same developmental shift occurring for the encoding of visual items is also evident in verbal memory tasks. Hasher and colleagues demonstrated that before the age of 8 years-old children remember better words semantically related to each other (Hasher and Clifton, 1974). In the call-name condition of this study, animal calls and names have different acoustic features. Thus, 6-7 year-olds could not rely on the acoustic features of the stimuli when asked to pair them. On the other hand, older peers may have memorized the words by paying attention to the concept behind them (in this case, the corresponding animal) or by using the subvocal rehearsal to prevent their decay in memory (Hulme and Tordoff, 1989). These findings confirm our initial hypothesis by indicating that the phonological loop affects how spatialized sounds are remembered and associated. The development of WM functionality (i.e., the phonological loop) affects audio-spatial memory abilities in sighted children. The changes in words' encoding, and the more use of the subvocal rehearsal influences how the sounds displaced on *ARENA2D* are stored and associated.

Finally, thanks to this paradigm, we have been able to understand how children, based on their age, explore a complex auditory structure. In this test, participants were asked to localize, remember, and eventually pair sounds displaced randomly on *ARENA2D*. Most of the previous research demonstrated that in the visual domain, children start integrating metrical and categorical representations of space at 4 years of age (Blades and Spencer, 1994; Newcombe, 2002; Sandberg et al., 1996).

Moreover, before the age of 7 years-old, they are not able to properly integrate spatial metrics, at least for the visual modality (Vasilyeva and Lourenco, 2012). The current research reveals similar difficulties when children explore a complex audio-spatial structure, in line with the starting hypothesis. Indeed, regardless of the experimental condition, 6-7 year-olds are more prone to start consecutive attempts by touching the same stimulus location more often than the older participants. The audio-anchor indicates that before the age of 8 years-old, children do not account for the mutual relations between the sounds, and, as a consequence, they do not adequately integrate the spatial positions with nearby items. Older participants instead show the opposite. The use of mutual relations permits them to hold and accurately process the information stored in memory to generate a functional representation of sounds' dispositions. This responds to the second scientific question of this study. Due to the changes in spatial skills during the first period of life (i.e., before the 8th year of life), children do not adequately construct a unified representation of the locations occupied by the sounds. Conversely, they need to refer all the locations to a fixed position. Thus, there is a development trend in the processing of spatial information also in the auditory domain, along with the ability to explore sophisticated audio-spatial landmarks.

The test was designed in the form of a game to be easily accomplished by children and young individuals. This procedure has the potential to inspire future protocols aimed at strengthening descriptive and categorical associations between concepts and their meanings, especially in sensory and cognitive disabilities. Audio-spatial memory skills can be trained in case of early visual deprivation using experimental procedures like the one presented in the current research. Finally, this study's findings can act as a launchpad for studying brain mechanisms underlying cognitive and spatial abilities through the auditory channel.

3.2 The effect of blindness on audio-spatial memory abilities

Since we observed the efficacy of the *Audio-Memory* on typically developing children (see section 3.1.1), we decided to use the same procedure to test audio-spatial memory skills in congenitally blind individuals. One of the research topics investigated in these years concerns the study of how the lack of visual experience affects high-level cognitive abilities (e.g., spatial memory and imagery). This paradigm is suitable for these purposes as it allows to create acoustic spatial layouts to be learned and processed. Very little is known on how blind (and sighted) individuals process a complex spatial structure in the auditory domain. The influence of blindness in the storing and sequential processing of sounds has been rarely studied (Setti et al., 2018; Vuontela et al., 2003) and, to our best knowledge, no one has investigated how blind individuals construct and manipulate a "dynamic" auditory structure. With the term "dynamic," we refer to a spatial structure whose configuration changes during the execution of the task. To fill this gap, we studied the capacity of blind people to hold in memory the locations of spatialized sounds (animal calls). However, the test also required the participants to update the mental representation of stimuli' positions continuously. We used the rules of the *Audio-Memory* described in the previous section (see section 3.1.1). Also, in this case, participants were required to match animal calls displaced on the surface of *ARENA2D*. When a new sound was discovered, or two sounds were paired, the experimenter covered the free slots with cardboard grids. The participant had to remember all these changes to end the test as soon as possible. In this study, all the stimuli were animal calls because we were not interested in studying the influence of the phonological loop. Moreover, to investigate the impact of the memory load on performance, we designed two experimental conditions by increasing the number of sounds to be paired: 4

couples in the first condition and 12 couples in the second condition (see figure 3.6). With this *Audio-Memory*, we addressed the following research questions:

1. To what extent does early visual deprivation influence audio-spatial memory skills?
2. What is the exploration strategy used by the two groups when asked to explore a complex auditory structure?

We hypothesized that, because of the lack of visual experience, congenitally blind participants would explore differently *ARENA2D* from the other group, since still anchored to sequential processing of space. Furthermore, due to the influence of the cognitive load and the difficulties in using the spatial relations among the sounds, they should be outperformed by the sighted.

3.2.1 Material and Methods

Sample

Twelve congenitally blind (9 females, 4 adults, mean age ± SD: 25.5 ± 15.29 y.o.) and twelve sighted (9 females, 6 adults, mean age ± SD: 25.83 ± 15.86 y.o) individuals took part in the experiment. Clinical details relative to their visual impairment are shown in Table 3.2. Blind adults were recruited from our institute database, while blind adolescents from the "Istituto David Chiossone," based in Genoa, Italy. The ethics committee approved the experiments of the local health service (Comitato Etico, ASL 3, Genoa, Italy), and parental or adult informed written consent for the study was obtained in all cases. All experiments were performed following the declaration of Helsinki. None of the sighted nor the congenitally blind participants had additional sensory disabilities. The test did not have a fixed duration since it depended on the ability of the participant. However, it lasted 25 minutes on average.

Experimental Procedure

The experimental protocol is an adapted audio version of the classical card game, *Memory*, to be performed by blind individuals (similar to the paradigm presented in section 3.1). The cards were replaced with sounds (animal calls) spatially displaced. The test, carried out with *ARENA2D* (see section 2.1), was divided into two experimental conditions depending on the cardboard grids placed on the device (see Figure 3.6). The grids differed in the number of apertures and their shapes. In the first condition, the apertures were as big as the haptic blocks, while in the second, they were smaller (4x4 cm). In the first condition, participants searched for 4 pairs of identical animal calls; on the contrary, in the second condition, the

Table 3.2 Participants' clinical details

PARTICIPANT	GENDER	AGE	RESIDUAL VISION	PATHOLOGY
S1	M	14	Uveitis	Lights and Shadows
S2	F	13	Retinopathy	Lights and Shadows
S3	F	12	Retinopathy	No Vision
S4	F	15	Laber's Amauroswe	No Vision
S5	F	15	Cataract	No Vision
S6	M	52	Retinopathy	No Vision
S7	F	30	Retinopathy	Lights and Shadows
S8	F	12	Glaucoma	No Vision
S9	F	42	Laber's Amauroswe	No Vision
S10	M	25	Retinis Pigmentosa	No Vision
S11	F	52	Retinis Pigmentosa	No Vision
S12	F	24	Retinis Pigmentosae	No Vision

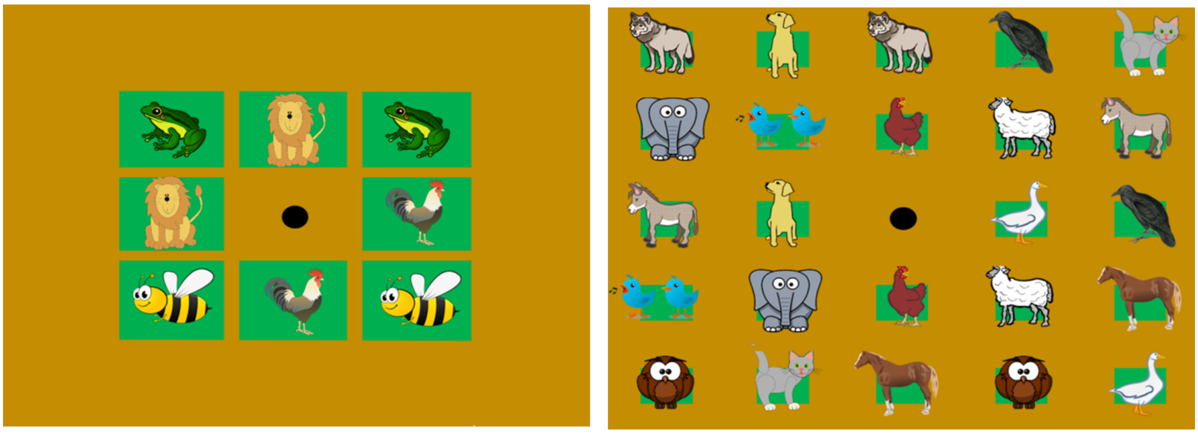


Figure 3.6 **Grid Used in the experimental conditions.** The two grids differ in the size of the apertures for each auditory stimulus. The apertures on the grids represented in the left column were 10 cm x 10 cm, equal to the haptic block size. The apertures on the grids represented in the left column were 4 cm x 4 cm. Depicted animals placed inside the squares, refer to the positions of the animal calls in both grids (images downloaded from a royalty-free website, https://publicdomainvectors.org/). The black dot at the center indicates the speaker emitting the feedback sounds.

pairs were 12. These grids have been designed in collaboration with rehabilitators of the "David Chiossone" Institute, a center for blind and low vision children based in Genoa. We used two different sets of animal calls for the two experimental conditions. In the second condition, the apertures were smaller because, as rehabilitators suggested, this size would have been more suitable for blind individuals to haptically explore the device and code the position of each stimulus. The second condition requires an increased amount of memory load. The central speaker to emit the feedback sounds, which is a “TADA” when two twin sounds were paired or a recorded voice saying “NO” in the other case. During the experiment, the subjects sat on a chair at a fixed distance of 30 cm from *ARENA2D*. The device position was adjusted to align the subject’s nose with the central aperture of the grid. None of the subjects had ever interacted with the array of speakers, and the group of sighted entered the room blindfolded. The test started with the first experimental condition (see Figure 3.6, panel on the left). The grid with 8 apertures was placed on the surface of *ARENA2D*. The experimenter guided the subject’s hands to explore the grid and counted with the participants the free slots by guiding their hands over the grid and its apertures. After this phase, the subject freely touched *ARENA2D* with both hands to familiarize themselves with the device. In this test, the subjects used only the index finger of the dominant hand. Before starting the experiment, they practiced with a trial session. With the 8-slots grid used in the first condition, the experimenter guided the participant’s hand over two free slots, first unpaired (that emitted different sounds), and then paired to make the participant familiar with the feedback sounds. After this practice session, the subjects listened to and identified the animal calls one by one. Thus, the participants performed the first experimental condition. Once this session finished, the second grid was placed over the device. The subjects explored the grid, first by counting the free apertures with the help of the experimenter, and then they touched the device with no guidance. When the subjects were confident with the grid, the second condition started. Two jingles were reproduced from the central speaker to make the test amusing at the end of each condition. Each song lasted 20 seconds. The sound pressure level (SPL) was maintained at 70 dB. Their RMS (Root Mean Square) level was calibrated to be the same across the various signals.

Data analysis

In this audio memory, to quantify subjects’ performance, we evaluated the memory score (see Figure 3.2) for both the first and second experimental conditions, the number of attempts, and the audio-anchor (see Figure 3.3). These were the same parameters used in the *Audio-Memory* for children (see section 3.1). The number of attempts and the audio-anchor were evaluated only for the second condition.

3.2.2 Statistical analyses

All the analyses were carried out in RStudio (RStudio, version 1.1.463). The hypothesis of normality was checked with the Shapiro-Wilk test (*shapiro.test()* R function). Analysis of the score, was carried out separately for the two experimental conditions, and the hypothesis of normality was confirmed only for the first one. The number of attempts and the audio-anchor instead, were evaluated only for the second condition. In order to check the influence of age on the overall performances, we ran for the three parameters, 2-ways (2x2) ANCOVAs (*aov()* and *aovp()* R functions in case of data normally distributed or not respectively) with *Group* (either blind or sighted) as between factor and *Age* (i.e., the age of the participants) as covariate for both conditions. The *aovp()* function is included in the *lperm()* package. The latter uses permutation tests to obtain p-values for linear models when the data do not follow a normal distribution. Post-hoc analyses were run with two-tailed unpaired Student's t-tests in case of data normally distributed (*t.test()* R function). Otherwise we used the t-test based on permutations ((*perm.t.test()* R function)). Cohen's value was determined to quantify the effect size (Cohen's d).

3.2.3 Main Results

With regard to the score, normality was not confirmed only for the first condition. The analysis of the first condition revealed only a significant main effect of the *Group* (number of repetitions=5000, p<0.01, Cohen's d=0.63), no significant main effect of the *Age* (number of repetitions=51, p=0.9, Cohen's d=0.09) nor significant interaction *Age*Group* (number of repetitions=51, p=0.82, Cohen's d=0.32). The same results were confirmed for the second condition: a significant main effect of the *Group* (F(1,20)=9.81, p<0.01, Cohen's d=0.71), no significant main effect of the *Age* (F(1,20)=0.11, p=0.74, Cohen's d=0.33) nor significant interaction *Age*Group* (F(1,20)=3.54, p=0.074, Cohen's d=0.38). We concluded that the age of the participants did not influence the pattern of results and the analyses (for the score) were carried out by putting together young individuals and adults. We observed that, regardless of the experimental conditions, the group of sighted reached a higher score in both the *First* (t-value=2.49, p<0.01, df=17.85, Cohen's d=1.02) and *Second* conditions (t-value=3.02, p<0.01, df=22, Cohen's d=1.2). This indicates that blind participants returned more times on the same speakers (see Figures 3.7). Regards the number of attempts, the results pointed out that the number of attempts employed to pair the stimuli was not influenced by the age of the participants. Analysis of variance indeed, only revealed a main effect of the *Group* (F(1,20)=7.8, p<0.01, Cohen's d=0.81), and no significant effect of *Age* (F(1,20)=0.09, p=0.84, Cohen's d=0.1) nor interaction between the two factors (F(1,20)=0.18, p=0.53, Cohen's d=0.28). Post-Hoc analysis instead highlighted that blind subjects, on average, needed more attempts to pair the sounds once their positions were discovered on *ARENA2D*

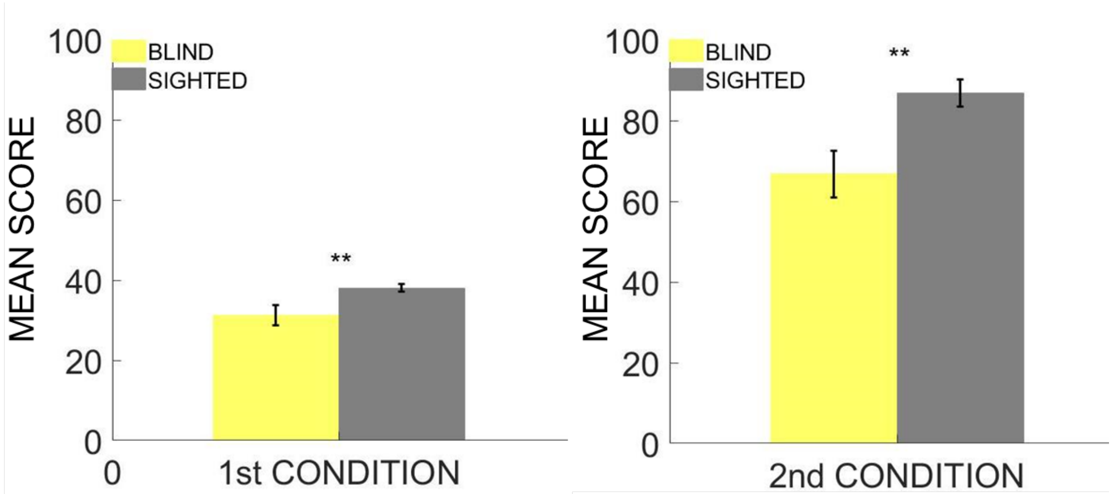


Figure 3.7 **Score.** Data are presented as mean and standard error. The panel on the left refers to the first condition, the panel on the right to the second. Regardless of the experimental condition, the group of blind reached a lower score. In the figure, the two asterisks (**) represent p<0.01.

(t-value=5.04, p=0.03, df=19.43, Cohen's d=0.57). Finally, we determined the effect of the age on the use of the audio-anchor and again we only found a main effect of the *Group* (F(1,20)=8.2 p<0.01, Cohen's d=0.92), and no significant effect of the *Age* (F(1,20)=0.63, p=0.72, Cohen's d=0.34), nor interaction between the two factors (F(1,20)=2.1, p=0.12, Cohen's d=0.23). We observed that in the second condition, blind participants were more prone to the use of this exploratory strategy to accomplish the task (t-value=3.78, p=0.041, df=20.12, Cohen's d=0.6).

3.2.4 Discussion

In the current study, we compared audio-spatial WM abilities in blind and sighted individuals with the *Audio-Memory* designed with *ARENA2D*. We examined the influence of early visual deprivation on audio-spatial memory abilities and on the strategy used to explore a complex acoustic structure. The results highlight that, in both experimental conditions, the group of sighted outperformed the blind group by reaching a higher score (see Figure 3.7). Furthermore, in the second experimental condition, the group of blind subjects needed more attempts to proceed with the task. They returned more times on the same speakers compared to the other group (see Figures 3.9 and 3.8). The first reason behind this pattern of results has to be ascribed to difficulties in combining the spatial positions of the sounds in a coherent and functional representation. The *Audio-Memory* required the active manipulation

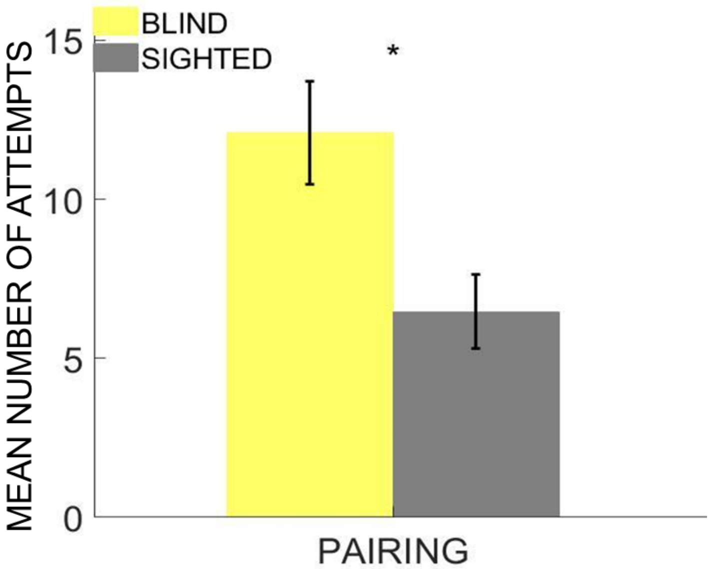


Figure 3.8 **Number of attempts** Data are presented as the mean across participants per each group and standard error. The sighted needed less attempts to pair the stimuli once their locations have been discovered on *ARENA2D*. In the figure, the asterisk (*) represents p<0.05.

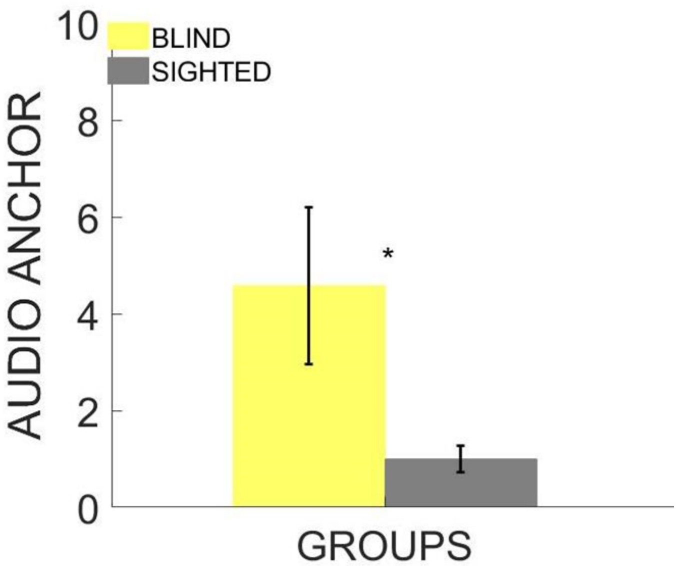


Figure 3.9 **Audio-Anchor.** Data are presented as the mean across participants per each group and standard error. The group of blind relied more on the use of the audio-anchor. In the figure, the asterisk (*) represents p<0.05.

of spatial information, thus affecting the performances of the blind participants. The subjects indeed, not only had to remember sounds' locations, but they were also asked to update this spatial representation. When two sounds were paired, their locations were covered by cardboard squares. Conversely, when a new sound was discovered, the participants had to remember its location and update the spatial representation by adding the new sounds in the spatial representation of the scene thus increasing the cognitive load imposed by the experimental paradigm. As already stated by Cornoldi and colleagues (Beni and Cornoldi, 1988), when a spatial WM task involves a great demand on memory, congenitally blind individuals experience more difficulties than sighted. After the exploration of a spatial layout in both haptic and auditory modalities, blind people can remember the positions of all the items or targets composing the configuration (Cornoldi et al., 2000; Vecchi, 1998; Vecchi et al., 1995, 2005). However, the *Audio-Memory* required the use of the spatial relations among the sounds composing the complex auditory structure. Early visual deprivation, at least in the haptic modality, specifically affects the simultaneous maintenance of multiple stimuli in memory. Vecchi and co-workers pointed out that blind people have difficulties in dealing with multiple stimuli at the same time when haptically conveyed (Vecchi et al., 1995). Similarly, our results show that, in comparison to sighted, blind individuals showed worse performances in organizing and maintaining spatialized auditory information. Likely, visual experience influences such ability. On the one hand, in sighted people, vision allows the simultaneous perception and manipulation of multiple stimuli at the same time. On the other hand, blind people need to rely on haptic and auditory modalities to process external stimuli, and such sensory modalities are based on sequential processing of space. Considering the calibration theory regarding the multisensory processing of spatial perception (Gori et al., 2008, 2010), the lack of vision since birth might have influenced the ability to process simultaneous spatial information. Congenitally blind participants tend to construct their mental representations piece-by-piece, and the building up and consequent processing of a unified image requires a high WM load. Thanks to the visual experience, sighted people code spatial information in the form of global, externally based representations (Cattaneo et al., 2008; Cornoldi et al., 1993; Pascual-Leone and Hamilton, 2001). Along this line, the group of blind participants tended to use the audio-anchor strategy to explore the device more often compared to the group of sighted (see Figure 3.5), thus indicating that they were more prone to construct their spatial representation piece-by-piece.

The results of this study suggest a different use of the mutual relations between the sounds (i.e., the frames of reference). The two main frames of reference used to represent the locations of entities in space are egocentric and allocentric (Ruggiero et al., 2012). The first defines locations of items in the surrounding environment with respect to the observer. Conversely, allocentric reference frames serve to encode spatial information by considering external landmarks. Pasqualotto and co-workers (Pasqualotto et al., 2013) stated that sighted individuals usually rely on allocentric frames of reference to orient themselves or to simply represent and process spatial information (Iachini et al., 2014). On the contrary, congenitally blind individuals make more use of egocentric frames of reference that are centered to the body (Thinus-Blanc and Gaunet, 1997). In the *Audio-Memory*, the lower use of the mutual relations among the sounds can be observed by the tendency to start consecutive attempts by touching the same speaker (i.e., the audio-anchor, see Figure 3.5). To construct a functional spatial representation of space, indeed, blind participants might not use the mutual relations among the sounds. They likely refer all the spatial locations to a position previously explored. In conclusion, in the current study, we evaluated spatial and memory skills in blind and sighted individuals along with the strategy of exploration of the complex auditory structure. As regards to the initial scientific questions, we found that early visual deprivation affects the processing and exploration of the sounds embedded in a complex acoustic structure. When the task demand is high, non-sighted subjects need more time to manipulate the spatial information learned during the task compared to the group of sighted. Furthermore, to build a functional unified representation, blind participants need to refer all the spatial locations to a unique position. In line with previous findings, we found that the limitations already seen in the haptic domain, are also valid for the auditory modality. The current paradigm, designed in the form of a game, can be used as a starting point to define novel procedures for both the clinical evaluation and rehabilitation of cognitive impairments. Through these solutions, the visually impaired children may quicken the learning and development of new concepts and associations, facilitating their inclusion in educational contexts.

3.3 Relationship between audio-spatial memory and spatial imagery

Previous research on the haptic modality demonstrated that early visual deprivation affects the ability to process spatial information held in memory (Aleman et al., 2001; Cattaneo et al., 2008). Even though visually impaired people can generate mental images of layouts through haptic exploration or verbal description (Kennedy, 1983), the ability to process such representations is strongly compromised, as seen by their performances in mental rotation and mental scanning tasks (Beni and Cornoldi, 1988; Vecchi et al., 2005). In these experimental procedures, participants are asked to rotate mentally (i.e., mental rotation) or to scan across a mental image (i.e., mental scanning) (Beni and Cornoldi, 1988). Although vision and touch are the most used sensory modalities for the evaluation of spatial memory, also audition conveys complex spatial information. However, only a few studies assessed the role of audition in spatial memory. In this study, we evaluated for the first time audio-spatial memory skills in blind and sighted individuals through complex auditory contents. To understand the role of visual experience in the development of auditory spatial representations, we designed an experimental paradigm by taking inspiration from the *Corsi-block* test. Differently from the original procedure, the blocks were replaced by sounds spatially arranged. The test was divided into two experimental conditions, based on the nature of the stimuli: non-semantic (pure tones plus white noise) and semantic (meaningful sounds). As shown in Figure 3.10, the semantic stimuli were arranged to recreate a coherent scene. During the experimental procedure, the participants learned these complex acoustic layouts and used them for later recalling. With this paradigm, we verified whether a coherent semantic elaboration could enhance memorization processes. The main research questions we wanted to address were:

1. What is the effect of blindness on audio-spatial memory abilities?
2. Does early visual deprivation influence the ability to construct an accurate representation of a complex auditory spatial layout?
3. To which extent does the knowledge of a coherent acoustic scene influence the performance in an audio-spatial memory task, especially in case of early visual deprivation?
4. Does the nature of the stimuli (i.e., non-semantic vs. semantic) affect memory performances?

Previous works showed that blind individuals can memorize a haptic spatial layout but show limits when asked to manipulate this spatial representation (Cornoldi et al., 2000; Cornoldi and Vecchi, 2004; Vecchi et al., 2005). Based on these results, we hypothesized that without a priori knowledge of sounds' dispositions, both groups should perform equally. However, after the learning of the semantic spatial layout, congenitally blind participants would show lower performances due to the request to use the mutual relations among the sounds and to the influence of the cognitive load in the recalling phase. The *ARENA2D* was used for this experimental procedure because it allows the spatial emission of sounds and the design of complex auditory structures. Furthermore, this audio-tactile tablet permits to register touch positions and to quantify the accuracy in the responses (i.e., the distance between the speaker that emitted the sound and the perceived position).

This study (https://www.nature.com/articles/s41598-018-31588-y) was published as peer-reviewed article.

3.3.1 Material and Methods

Setup and Stimuli

The sounds were reproduced through *Arena2D* (see Figure 2.1 and Chapter 2 for details). We used 13 sounds for each experimental condition, each lasting 3 seconds. The non-semantic sounds consisted of pure tones varying in frequency from 250 to 1300Hz, with a step size of 50Hz. White noise was added to the tones to facilitate their localization. All the sounds in the non-semantic condition were created in Matlab (R2013a, The MathWorks, USA). In the semantic condition, meaningful sounds were used (e.g., wind, bee, sheep). They were downloaded from a royalty-free web archive (Freesound.org). Each stimulus was positioned to a specific haptic block to recreate a coherent scene (we chose a countryside environment, see Figure (3.10)). The sound pressure level (SPL) was maintained at 70 dB, and the RMS (Root Mean Square) level was calibrated to be the same across the various signals. To preserve a correspondence, in both experimental conditions, the speakers that reproduced the same sounds were equal (e.g., the first and the second block in the semantic condition, also emitted the same non-semantic sounds).

Experimental Procedure

The experimenter positioned the participants at a fixed distance of 30 cm from the center of *ARENA2D*. In the listening phase, they placed the index finger of the dominant hand on a plastic lid serving as a reference point. None of the sighted subjects had seen the device beforehand, and all of them entered the experimental room blindfolded. They familiarized themselves with *ARENA2D* by haptically exploring the device and counting together with the experimenter the number of speakers. The participants were trained with a practice session only with non-semantic sounds. The test was divided into pre- and post-test phases, punctuated by an exploration. Subjects listened to sequences of sounds of increasing length. They had to recall the sequences in the same order of presentation by touching the positions from which the sounds were emitted. Auditory feedback indicating response acquisition was provided to allow the participants to release their finger from the touched position. To make the test amusing for the subject, we employed a cat sound. The feedback sound did not interfere with the overall performance (i.e., it is not a distractor). The participants heard the sound since the trial session and were instructed to ignore it during sequences recalling.

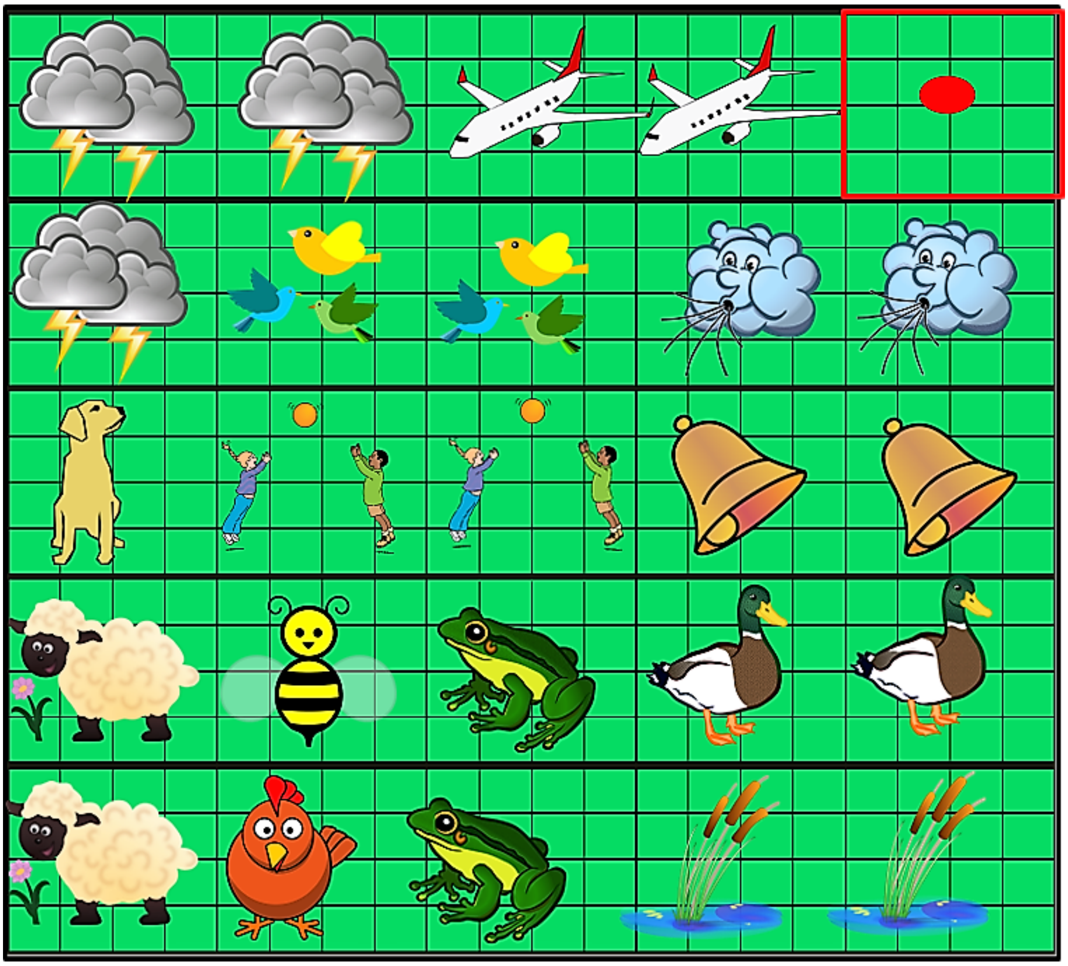


Figure 3.10 **Semantic Auditory Scene**. The square highlighted in red at the right corner of the picture represents a haptic block (the red dot in the center is the speaker, see section 2.1 for details). Each square is composed of 16 smaller squares (each representing a tactile sensor). The pictures represented on *ARENA2D* surface are the sounds used in the semantic condition. Each picture represents the sound emitted by the specific block. The sounds related to the sky (e.g., the plane, the wind) are placed on the top of the device. Sounds related to the ground are at the center, while sounds related to nature (animals and pond) are emitted from the bottom loudspeakers. All the images in the picture have been downloaded by a royalty-free images web archive (https://publicdomainvectors.org/).

After each sequence, the participants placed the index finger back on the reference point, while waiting for the next trial. A time interval of 1s separated two consecutive sequences, and the change in the length was verbally communicated by the experimenter (varied among 2, 3, and 4 sounds). We used two different sequences per length, for a total of 6 trials per block. Differently from the original paradigm, the increase in length did not depend on subjects' performance, but the participants were tested with all the sequences (n=6). For

each trial, the spatial distance between two consecutive stimuli was set at 20 cm to facilitate the discrimination among the sounds. The aim was to test subjects' performances before and after the exploration phase with sequences of different lengths. Before starting the semantic condition, all the participants listened to the meaningful sounds one by one. The experimenter asked to identify each of them. In the exploration phase instead (5 minutes lasting), participants freely touched *ARENA2D*, while listening to the sounds coming from the speakers and trying to imagine the scene created with their combination. They explored the device with the index finger of the dominant hand, and the sound started to be emitted soon after the participant touched a tactile sensor.
The participants were also instructed to pay attention and remember the position of the sounds after each touch and, in the semantic condition, to try to learn the semantic acoustic spatial layout. However, the experimenter told them that the scene was coherent and that the stimuli were not randomly displaced. Finally, at the end of the semantic condition, the subjects rebuilt the scene by indicating the positions of the single items composing it. The experimenter read aloud the names of the sounds, and, after each name, the subject touched the corresponding location on *ARENA2D*. The performance was quantified in terms of percentage of single sounds (n=18) and sequences (n=6) correctly recalled and the memory span. Following a validated procedure (Röser et al., 2016), the span was calculated as the product between the number of correctly recalled trials and the sequence lengths, summed up and divided by 2 (see equation 3.1). A single response was considered correct if the subject touched *ARENA2D* at a distance less than 14 cm (distance between the centres of two consecutives speakers in the diagonal direction) from the position of the loudspeaker that emitted the sound.

$$\sum_{i=1}^{4} \frac{corrsubtrials(i) * i}{2} \tag{3.1}$$

In the formula 3.1, i is the sequence length and *corrsubtrials* is the number of trials recalled for the i-th length. According to 3.1, the span is between 1 and 10. In span's evaluation we started with sequences of 2 sounds, assuming the trials of 1 sound as correct.

Sample

Eleven congenitally blind (7 females, mean age ± SD: 41.82 ± 14.6070) and 11 sighted (7 females, mean age ± SD: 38.8182 ± 11.3033) individuals took part in the experiment. In the sighted group, all had normal visual acuity. Table 3.3 refers to the information about blind subjects, that were recruited from the Italian Institute of Technology (IIT) database and did not report additional sensory or cognitive impairment. The study was approved by

Table 3.3 Participants' clinical details

PARTICIPANT	GENDER	AGE	PATHOLOGY	RESIDUAL VISION
S1	M	58	Glaucoma	No Vision
S2	M	57	Uvetis	Lights and Shadows
S3	M	51	Retinopathy	No Vision
S4	M	25	Laber's Amauroswe	No Vision
S5	F	30	Retinis Pigmentosa	No Vision
S6	F	41	Glaucoma	No Vision
S7	F	52	Retinis Pigmentosa	Lights and Shadows
S8	F	27	Retinopathy	No Vision
S9	F	62	Atrophy Eyeball	No Vision
S10	F	30	Retinopathy	No Vision
S11	F	30	Microphthalmia	No Vision

the ethics committee of the local health service (Comitato Etico, ASL 3, Genova, Italy) and conducted in accordance with the Declaration of Helsinki (1964). Informed written consent was obtained for all the participants and the experimental procedure lasted for 30 minutes.

3.3.2 Statistical analyses

For the evaluation of the improvement after the exploration, we carried out a 2-way (2×2) mixed-measures model ANOVA with *Group* (either blind or sighted) and *Condition* (either semantic or non-semantic) as between and within factors respectively. The improvement is defined as the difference between the score in the post- and pre- exploration phases. Therefore, we also analyzed the scores in the pre- and post- exploration phases separately by performing a three-way (3×2) mixed-measures model ANOVA. The model consisted of the same factors described above (*Group* and *Condition*) and *Phase* (either pre and post) as within factor. Post-hoc analyses were carried out with two-tailed Student's tests, both paired and unpaired. *Bonferroni* correction was used to test the significance of multiple comparison Post-Hoc tests ($p<0.05$ was considered significant). All the analyses were run in Matlab (Matlab R2017a, The Mathworks). The data are presented in the form of bar plots in terms of mean and standard error. Cohen's coefficient was evaluated to quantify the effect size (Cohen's d).

3.3.3 Main results

For the analyses of performance, we first considered the memory improvement, defined as the difference between post- and pre-test phases. In Figure 3.11, results are shown for the two groups. The analysis of variance revealed significant interactions between *Group* and *Condition* for single sound, ($F(1,20)=15.49$, $p<0.001$, Cohen's d=0.88), span ($F(1,20)=14.45$, $p<0.01$, Cohen's d=0.83) and correct sequence improvements ($F(1,20) =14.46$, $p<0.01$, Cohen's d=0.73). Only the group of sighted reached a greater improvement in the semantic condition compared to the non-semantic (paired t-test, single sound improvement: t= 3.83, $p<0.001$, df= 20, Cohen's d= 1.78; correct sequence improvement: t=3.85, p=0.013, df=20, Cohen's d= 1.46; span improvement: t=5.39,p=0.013, df=20, Cohen's d=1.42), while for the blind we did not find any significant difference between the improvements in the two experimental conditions (paired t-test, single sound improvement: t=5.02, p=0.67, df=10, Cohen's d= 0.23; correct sequence improvement: t=4.01, p=0.99, df=10, Cohen's d= 0.25; span improvement: t =4.97, p =0.87, df=10, Cohen's d=0.18). Moreover, the sighted participants benefited more than the blind from the exploration in the semantic condition (unpaired t-test, single sound improvement: t=3.8, $p<0.01$, df=20, Cohen's d=1.2; correct sequence improvement: t =3.33, p =0.012, df=20 Cohen's d=0.92; span improvement: t=3.71, p =0.022, df=20, Cohen's d=0.78). To see whether this outcome was due to a better ability in memorizing the auditory scene, we asked at the end of the semantic condition to indicate the locations of the single stimuli. The results did not reveal any significant difference in the items recalled by the two groups (unpaired Student's t-test: t=5.3, p=0.76, df=20, Cohen's d=0.19). Therefore, the reason why the sighted reach a greater improvement is not due to a better ability in remembering the auditory scene (see Figure 3.12). In order to verify whether the blind reached a ceiling level already in the pre-test phase, we analyzed separately the pre- and post-test phases. Results are shown in Figure (3.13). We ran 3-ways (3x2) mixed measures ANOVA, separately for the number of single sounds, the number of correct sequences recalled, and the memory span. Analyses of variance for the three parameters revealed significant interactions between *Group*, *Condition* and *Phase* for single sounds ($F(1,20)=17.99$, $p<0.001$, Cohen's d=0.92) correct sequences ($F(1,20)=14.46$, $p<0.01$, Cohen's d=0.74) and span ($F(1,20)=14.45$, $p<0.01$, Cohen's d= 0.81). Post-Hoc analyses revealed, in the semantic condition (see Figure (3.13)), that the group of sighted outperformed in the post-test phase compared to the blind (unpaired t-test, percentage of single sounds: t=4.08, $p<0.01$, df=20, Cohen's d=0.92; percentage of correct sequences: t=3.63, p=0.025, df=20, Cohen's d=0.76) and to the pre-test phase (paired t-test, percentage of single sounds correctly recalled: t=6.09, $p<0.01$, df=10, Cohen's d=0.98; percentage of sequences correctly

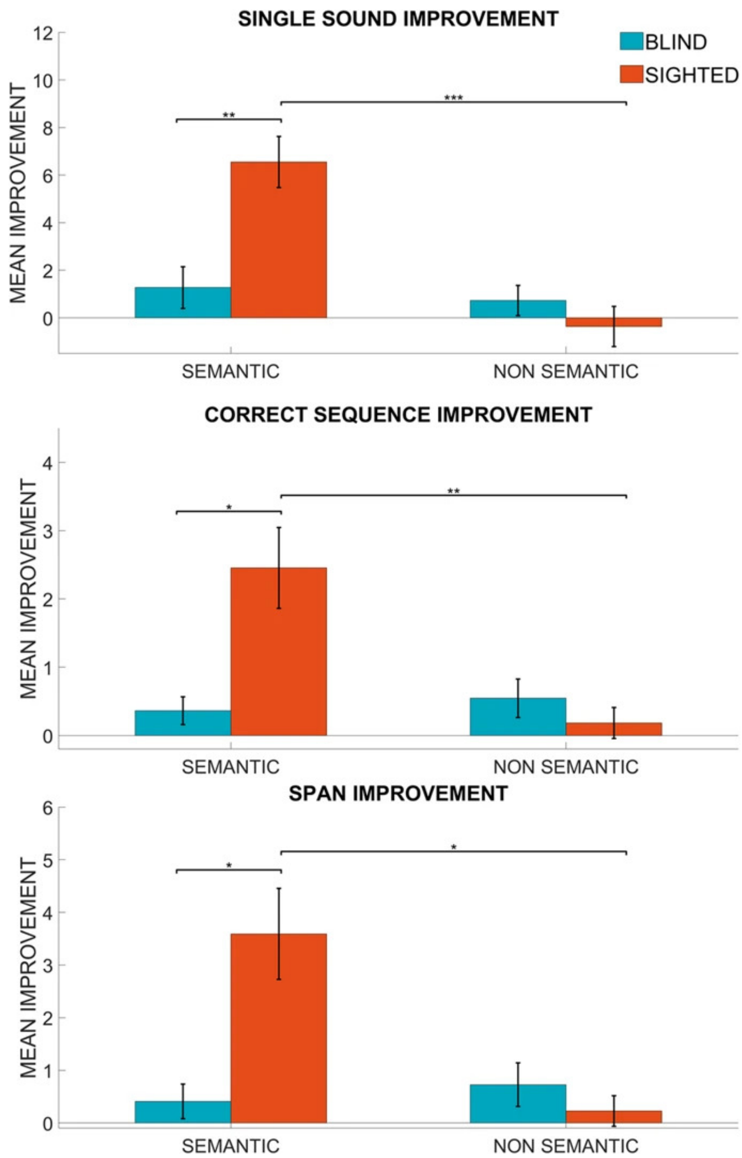


Figure 3.11 **Improvement of the two groups in both experimental conditions**. Data are presented in terms of mean and standard error. The bars represent the difference between the scores in the post- and pre-test phases for the number of single sounds correctly recalled (upper panel), the number of sequences correctly recalled (middle panel), and the span reached (lower panel). The improvement in the semantic condition was greater for the group of sighted compared to the blind and the non-semantic condition. There was no significant difference in improvement for blind participants between semantic and non-semantic conditions. The asterisks indicate statistical significance (one asterisk (*) represents $p<0.05$, two asterisks (**) represent $p<0.01$ and three asterisks (***) represent $p<0.001$).

recalled: $t=4.13$, $p=0.032$, $df=10$, Cohen's $d=0.72$). In addition, they performed better in the semantic rather than non-semantic condition (paired t-test, percentage of single sounds:

Figure 3.12 **Mean percentage of audio items recalled out of the total.** Data are presented as mean and standard error. In the figure, the bar plots show the percentage of items correctly recalled after the end of the semantic condition.

t=6.05, p<0.01, df=10, Cohen's d=1.23; percentage of sequences correctly recalled: t=4.37, p=0.03, df=10, Cohen's d=0.65), but only in the post-test phase. Furthermore, in the semantic condition, sighted individuals reached a higher memory span compared to the blind (unpaired t-test: t=3.37, p=0.012, df=20, Cohen's d=0.82) and to the pre-test phase (paired t-test: t= 4.21, p = 0.032, df=10, Cohen's d=0.58). Furthermore, comparison across conditions highlighted that sighted subjects recalled longer sequences in the semantic rather than non-semantic condition after the exploration (paired t-test: t=3.90,p= 0.04, df=10, Cohen's d=0.5). To summarize, the blind did not benefit from the exploration phase, regardless of the experimental condition. The performance before the exploration was the same among participants, and both groups were able to generate a mental representation of the auditory scene.

3.3.4 Discussion

In this work, we compared audio-spatial WM and imagery abilities in sighted and congenitally blind individuals with *ARENA2D*. The task proposed is based on mental imagery, since the subjects relied on their ability to learn an acoustic spatial layout and to use the gained information for later recalling. By taking inspiration from the *Corsi-Block* test, both groups were asked to recall sequences of spatialized sounds in two experimental conditions. The latter differed on the basis of the stimuli used: non-semantic (white noise plus pure tones at different frequencies) and semantic (meaningful sounds, see Figure 3.10). Sounds' dispositions were discovered after a five minutes exploration after the pre-test phase. Before this phase, the participants did not know that the sounds belonged to a coherent spatial layout,

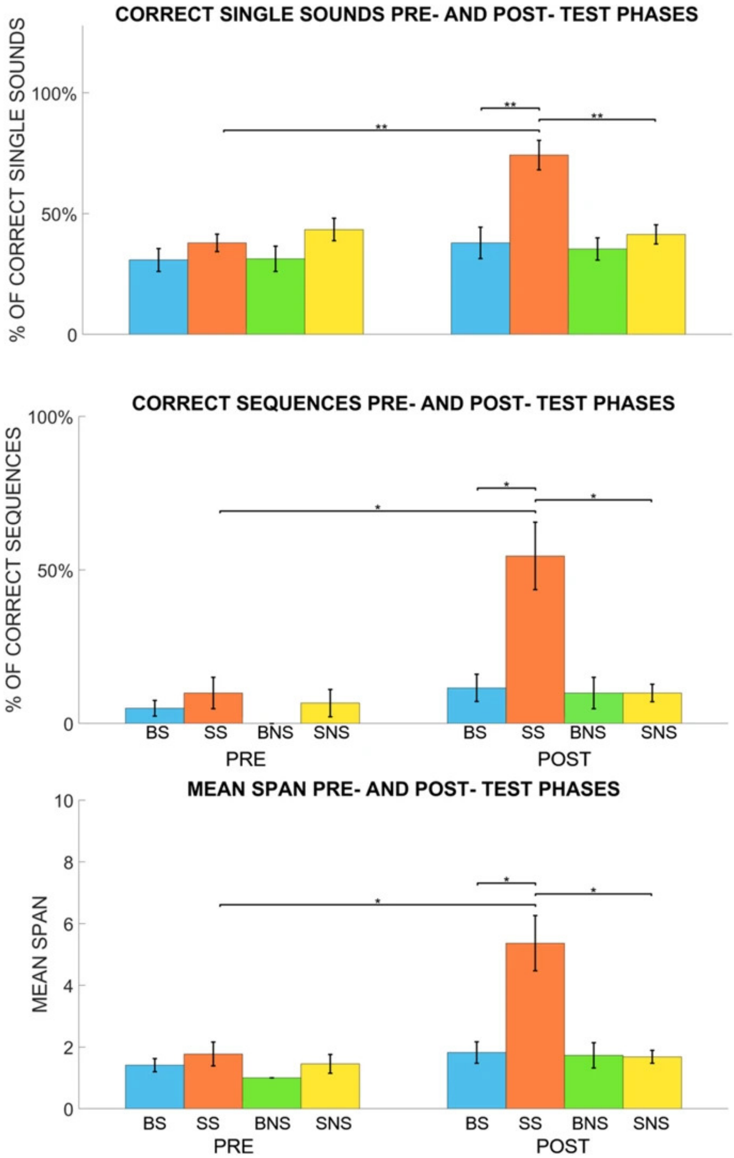


Figure 3.13 **Comparisons between pre- and post-test phases.** Data are presented as the mean and standard error. The labels BS, BNS, SS, and SNS refer to the experimental conditions and the groups (performance of the blind (B) and sighted (S) participants in the semantic (S) and non-semantic (NS) conditions). In the figure, the upper panel is the percentage of single items, the middle panel the percentage of sequences correctly recalled while the lower panel the memory span. Before the exploration, regardless of the experimental condition, the two groups performed similarly. Only the group of sighted, in the semantic condition, improved after the exploration, as indicated by the asterisks (one asterisk (*) represents p<0.05, two asterisks (**) represent p<0.01 and three asterisks (***) represent p<0.001)

.

especially in the semantic condition. The results highlight that before the exploration and regardless of the experimental condition, both groups performed similarly (in terms of single sounds, number of correct sequences recalled, and memory span, see Figure 3.13). Only the group of the sighted exhibited better performances after the semantic exploration (see Figure 3.11), and this was not due to a better ability in remembering the auditory items composing the acoustic scene (see Figure 3.12 for details).

We firstly investigated to what extent the semantic of the sounds influence the performance. Despite the semantic nature of the sounds, meaningful auditory stimuli are easier to be memorized than non-semantic ones when they are part of a coherent organized scene (Taevs et al., 2010). The semantic information per se does not lead to a benefit in memory performances (Mandler and Ritchey, 1977). Before the exploration, the participants focused more on the perceived origin of the sounds. Both groups were able to recall the positions of the single items composing the acoustic scene (Figure 3.12), in line with previous research (Cattaneo et al., 2008; Vecchi et al., 1995, 2005) and our starting hypotheses. We extended the conclusion made by Cattaneo and co-workers (Cattaneo et al., 2008) according to whom the representations generated after haptic explorations may be as accurate as of the ones based on visual imagery.

The improvement is present only after the exploration phase, supporting the hypothesis that only active manipulation of the retained information might meliorate memorization. The *Corsi-Block* test is a passive memory task since the only request is to remember sequences of spatialized sounds with no active manipulation of this information (Cornoldi and Vecchi, 2003). Conversely, in an active memory task, the participants are asked to process the incoming information (Vecchi et al., 2005). In the test here presented, the number of sequences to remember was low (from 2 to 4 sounds), and before the exploration, both groups performed similarly (Figure 3.13). Early visual deprivation did not influence the ability to perform a passive memory test, and audio-spatial abilities are similar in the two experimental groups. The difference in the performances only arises after the semantic exploration.

One reason for the difference in the pattern of results lies in the fact that blind subjects seem to have difficulties in combining the spatial locations of the auditory items in a coherent and functional representation, even when they are aware of the stimulus positions in the spatial configuration. After the semantic exploration, both groups understood and remembered that the sounds belonged to a coherent acoustic layout. However, only the sighted used this information for sequences' recalling. Thus, this study confirms what observed in the *Audio-Memory* (see Section 3.2). Auditory perception, sequentially in nature, affect the simultaneous perception and manipulation of several sounds. When a series was presented, the single sounds had to be combined to form the sequence to be recalled. Thanks to the visual experience, it is likely that the group of sighted made better use of the spatial relations among the sounds to combine them in a compact and functional representation.

Pasqualotto and co-workers (Pasqualotto et al., 2013) demonstrated that sighted individuals usually rely on allocentric frames of reference to orient themselves or to represent and process spatial information (Iachini et al., 2014). On the contrary, congenitally blind individuals make more use of egocentric frames of reference that are centered to the body (Thinus-Blanc and Gaunet, 1997). It is therefore likely that in the post-exploration phase, the group of sighted may have used the mutual relations among the sounds composing each series by relying on the object-to-object (allocentric) relationship between the stimuli, thus facilitating sequences recalling.
In conclusion, in the current work, we investigated how the learning of a complex audio structure influences the storing and processing of spatialized sounds. We demonstrated that early visual deprivation affects how acoustic spatial information is processed and used in a memory task. The results also confirm previous research in the haptic modality (Cornoldi et al., 2000; Cornoldi and Vecchi, 2004; Vecchi et al., 2005) because both groups were able to generate a mental representation of the acoustic structure. Still, only the sighted participants used this information in the subsequent recalling of the sequences. Therefore, in line with the starting hypotheses, we demonstrated that the semantic sounds were better memorized after the exploration, that is, after finding out that they belonged to a coherent spatial layout. Both groups were able to generate a mental representation of sounds' disposition, but only the sighted were able to use the learned spatial information. Only visual experience indeed permits the simultaneous processing of multiple inputs, and early visual deprivation affects the use of the spatial information in both the haptic and acoustic modality. Although the test here presented is not the exact replication of the *Corsi-Block* test, it can be nevertheless used to measure audio-spatial memory abilities in case of lack of vision. Furthermore, this paradigm can be employed to understand how blind people learn and memorize a complex acoustic spatial layout, as already profoundly discussed in the haptic domain (Vecchi et al., 1995, 2005).

Chapter 4

The *Audio-Corsi*: An acoustic VR adaptation of the *Corsi-Block* paradigm

In the visual modality, WM abilities are assessed with a wide range of procedures (Oberauer et al., 2005). In this context, the most used are span measures either verbal (Blankenship, 1938) or spatial (Berch et al., 1998). In these tests, participants are asked to maintain information in memory for later recall or immediate processing (St Clair-Thompson and Sykes, 2010). The memory span, used as a neuropsychological parameter for the identification of short-term spatial dysfunctions, measures the capacity of the VSSP of the WM model (Baddeley et al., 1986; Kessels et al., 2008). The span indicates how many items can be maintained in memory for a short time. One of the methods most used to quantify the memory span is the *Corsi-Block* test (Milner, 1971). The paradigm was administrated for the first time in 1973, as the visuo-spatial counterpart of the verbal memory span task (Milner, 1971). The *Corsi-block* test is used in both experimental and clinical settings (Kessels et al., 2000), especially for the diagnose of learning disabilities in children, dementia, and other neurological pathologies (e.g., Alzheimer, schizophrenia or Korsakoff's syndrome) (Carlesimo et al., 1994; Haxby et al., 1983). Additionally, the *Corsi-Block* test serves as an investigation tool for developmental changes and gender differences in spatial and memory abilities (Capitani et al., 1991; Orsini et al., 1986). In the context of experimental psychology, this task has also been useful to confirm the theoretical assumptions of the model proposed by Baddeley (Baddeley and Logie, 1999), such as the separation between the two slave subsystems (VSSP and phonological loop). Initially, it was thought that the *Corsi-Block* paradigm only acted on the resources of the VSSP (Smyth and Scholey, 1992). However, Vandierendock and co-workers (Vandierendonck et al., 2004) demonstrated that when the task demand requires the maintenance of information for recalling, new sources are required, especially for the longest sequences (Vandierendonck, 2000; Vandierendonck et al., 2004). The traditional apparatus consists of a set of 9 blocks randomly arranged on a wooden 23x28 cm board (Brunetti et al., 2014; Corsi, 1973) (see Figure 4.1). The examiner taps a series

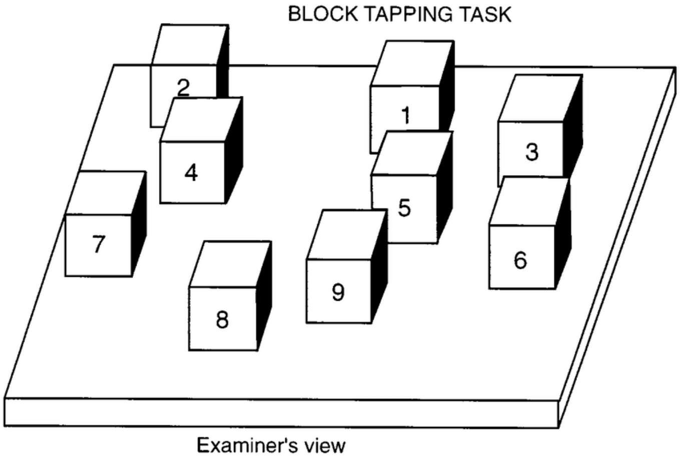


Figure 4.1 **Orginal *Corsi-Block* apparatus**. In the picture, the original illustration of the Corsi apparatus is shown. The drawing was reprinted from Corsi's doctoral dissertation in 1972. The blocks are numbered from 1 to 9 on the vertical surface to allow the experimenter to identify the blocks to be tapped

of blocks, and the subject attempts to reproduce it soon after the sequence's presentation. The participant recalls the sequences in the same order of presentation ("*forward* task") or in reverse ("*backward* task"). The number of trials per length usually varies among 3, 4, or 5. If a certain proportion of sequences for each length is correctly recalled, usually 1/2, 1/3, or 3/5, (Brunetti et al., 2014), then the length in the subsequent trials is increased by one. Otherwise, the test stops. Generally, to quantify the performance, three parameters are taken into account: the memory span (longest sequence correctly recalled), the total number of correct sequences, and the total score, which is the product between the two. Several computer based-forms of the *Corsi-Block* have been designed (Nelson et al., 2000). In these versions, the blocks are replaced by squares that flash on a computer screen. Thus, the subject reproduces the presented sequences by tapping the screen or using a mouse to click on the virtual blocks. Other versions of the *Corsi-Block* are the walking Corsi (Piccardi et al., 2008), the eCorsi (Brunetti et al., 2014), which is a software developed for the rendering of a tablet-based version of the test and the haptic Corsi proposed for the first time by Ruggiero and colleagues (Ruggiero and Iachini, 2010). In this last version, the position of the blocks in a sequence is encoded by fingers' movements. The spatial referencing is obtained by keeping participants' hands on the board. In this chapter, I am going to describe the adaptation of the

Corsi-Block test in the auditory domain, named *Audio-Corsi*. The test was carried out with the VR system described in section 2.2.

4.1 Design and development of the *Audio-Corsi* and validation of the system

In this study, visual and audio-spatial WM skills were compared employing the *Corsi-Block* test and our acoustic version, named *Audio-Corsi*. We wanted to study the differences and the similarities between the two paradigms, to shed light on the WM processing of spatial information for vision and audition. Similar to the classical *Corsi-Block* procedure, the subjects listened to sequences of spatialized sounds and recalled them using a custom-made keyboard (see Figure 2.10 in section 2.2.4). The *Audio-Corsi* was designed with a novel system based on an acoustic VR simulation. From a technological point of view, we verified whether this solution could be used as an investigation tool for the evaluation of spatial WM skills in the auditory domain. *ARENA2D* has several limitations, mainly because of its size and its cost. This device is hard to be reproduced, and it is not suitable for experimental and clinical research in very young individuals.
Furthermore, the use of real speakers very close to each other might influence the accuracy in localization, and, in case of a deterioration of sound quality due to extensive use, the substitution of the haptic blocks might require time and monetary efforts. Therefore, we designed the system based on acoustic VR, which is easily reproducible, cheap, and customized. This system might eventually open new ways for the early evaluation of cognitive skills (e.g., spatial memory), to intervene in the first period of life, especially in the case of congenital blindness. Here we present a validation study where the *Corsi-Block* test was adapted to the auditory modality. The main questions that led to the design of this study were:

1. Is it possible to study audio-spatial WM skills through an acoustic VR system?
2. Does the newly developed device permit to extract an equivalent audio-spatial memory span to quantify memory capacity in the auditory domain?

The acoustic VR apparatus had never been tested. With this study, by testing sighted individuals, we were interested in seeing the pros and cons of using such system in the context of audio-spatial memory. Since vision is the most accurate sense for spatial perception (Beni and Cornoldi, 1988), we expected a difference in WM performances across the two sensory modalities.

4.1.1 Task and Procedure

Sample

The current research involved 16 healthy participants (5 males, mean age $\pm$ SD: 24.75 $\pm$ 3.19). None of the subjects reported any sensory or cognitive impairment. Whereas they were informed about the procedures to be used, and the purpose of the experiment, the underlying research questions were not revealed to them. The ethics committee approved the experiments of the local health service (Comitato Etico, ASL 3, Genoa, Italy). Adult informed written consent for the study was obtained in all cases, and all the experiments were carried out following the declaration of Helsinki. The whole session did not have a fixed duration since it depended on the ability of the participant. However, it lasted ~25 minutes on average.

Corsi-Block test

The subjects sat in front of the experimenter with the board in between them. We used a wooden board like the one used by Kessels (Kessels and Postma, 2018; Kessels et al., 2000), see Figure 4.2). The board and the blocks were brown, and a number ranging from

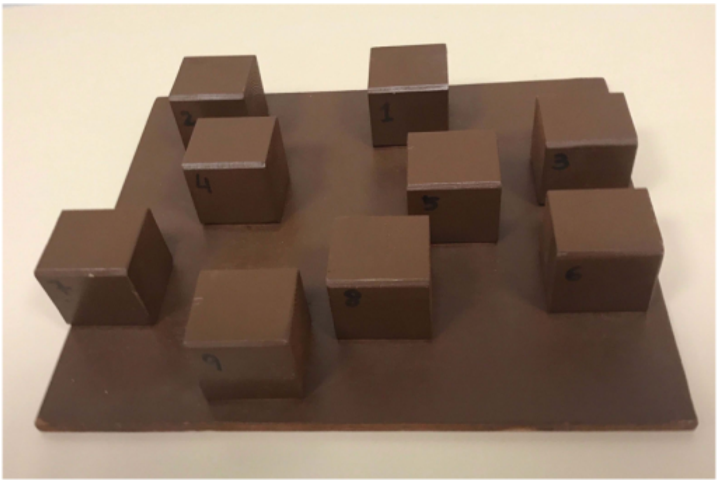


Figure 4.2 **Wooden Board used to accomplish the Corsi-Block paradigm.**

1 to 9 was printed on the side of the block directed towards the experimenter to foster the administration procedure. Participants were instructed to tap the blocks in the same serial order of presentation (3 sequences per each length from 1 to 9 blocks). The experimenter explained them that if at least one out of 3 sequences was correctly recalled, the test went on. During the presentation of the sequence, the experimenter tapped the blocks at a rate of one

block per second. The experimental session stopped when the participant did not correctly recall any of the 3 presented sequences. The sequences administrated in the *Corsi-Block* paradigm are the same used in clinical settings, according to normative standards (Spinnler, 1987).

Audio-Corsi

After the *Corsi-Block*, the subjects were blindfolded, and the experimenter replaced the set of blocks with the keyboard of the acoustic VR system (see section 2.2 for the technological details). At the beginning of the auditory condition, participants were asked to haptically explore the keyboard (Figure 2.10). When they felt confident with the task, the experimenter explained the disposition of the sources in the virtual environment and the test procedure. The stimulus used in the *Audio-Corsi* was pink noise (time length: 2.5 seconds). The reasons why we used pink noise for this validation study are explained in section 2.2.2. The auditory stimuli were presented binaurally through headphones from the locations shown in Figure 2.8. The right and left positions, closer to participants' head, were simulated by an interaural intensity difference of 20 dB (about 55 and 75 dB sound pressure level for each ear). The north-east and north-west locations were simulated by an interaural intensity difference of 10 dB (about 34 and 44 dB sound pressure level for each ear). For the frontal position, the noise was presented binaurally at an equal intensity (about 80 dB sound pressure level). Finally, regards the location from the back, since it was not perfectly aligned to the midline, we set an intensity difference of 8 dB (about 32 and 40 dB sound pressure level for each ear). Before starting the test, participants performed a 3-minute training to familiarize themselves with the use of the device and the virtual environment. The training served to let the participant know the virtual environment and how the sounds were reproduced from the locations. Even though the participants in the chosen configuration well-localized the sources, in the pilot studies, the north location was not well identified as the other five (see section 2.2.2). Therefore we introduced the training to be sure that the localization of the sources would have been optimal. Without this preliminary phase, there could have been errors in sequence's recalling not due to memory but to perceptual mistakes. In the *Corsi-Block*, the participants could see how the blocks were displaced; thanks to this training, they could understand how the sounds were spatially arranged. The subjects explored the keyboard while pressing the red buttons and listening to the sounds. They were asked to identify each sound position by pressing the corresponding button on the keyboard to verify if the training was effective. This procedure was done for each sound presented individually. Then the test started, following the rules of the *Corsi-Block* test. The subjects listened to sequences of sounds of increasing length, from

2 to 9 sounds, three sequences per length. During the listening phase, the participants placed the index finger of the dominant hand on the reference button (blue button, see Figure 2.10 in section 2.2.4). In the recalling phase, they pressed the buttons with the same finger following the order of the presentation of the sound stimuli. Afterward, they moved the index finger back on the reference point while waiting for the following sequence.

4.1.2 Data Analysis

According to the *Corsi-Block* test (Kessels et al., 2000), the parameters evaluated in both the visual and auditory tasks were:

- The memory span;
- The number of sequences to end the task;
- The product between the two parameters mentioned above (total score);

The memory span corresponds to the length of the longest sequence correctly recalled. The second parameter refers to the number of correctly recalled sequences out of the total (n=24). Finally, the product between the two was also calculated (the total score) (Kessels et al., 2008, 2000). The total score is more reliable compared to the memory span alone since it takes into account the performance of all the trials of a given length.
We calculated for each of the three parameters the coefficient of variation (CV), that is, the ratio between the standard deviation (SD) and the mean of the sample; the lower the value of CV, the more precise the estimate. In detail, distributions with a coefficient of variation less than 1 are considered to be low-variance, whereas those with a CV greater than one are considered to be high variance. Thus, the lower the CV, the closer the points are to the average. The CVs also served as a way to compare the variability in the scores (e.g., memory span) between the *Corsi-Block* and the *Audio-Corsi*.
We were interested in seeing whether the variability in the memory span reached in the two conditions was, on average, the same across participants. The evaluation of data variability served indeed to assess the validity of the auditory VR system. If in both tests, the performances exhibited by the subjects are similar (i.e., data dispersion is similar) and close to the average value, it is possible to infer that this system might be suitable for the definition of an audio-spatial memory span. Validation studies indeed identified, by means of the *Corsi-Block* test, a visuo-spatial memory span with a value ranging between 6 and 7 (Kessels et al., 2000), which is the standard range in the absence of cognitive delays. In other words, healthy subjects, on average, are able to maintain in memory a number of items up to

6 or at maximum 7, with low variability. We checked if, in the *Audio-Corsi*, the participants recalled the sequences up to a certain length with the same variability of the *Corsi-Block*.

4.1.3 Statistical analyses

Statistical analyses were carried out in Matlab (R2017b). Firstly, the normality of the data was checked with the Jarque-Bera test (*jbtest()* Matlab function). Then, the parameters listed above were compared between the two experimental conditions by running paired Student's t-tests (*ttest()* Matlab function). We also compared the variances in the data sets (Barlett's test *vartestn()* Matlab function). Cohen's value was determined to quantify the effect size (Cohen's d).

4.1.4 Results

The performances were compared between the *Corsi-Block* test and the *Audio-Corsi* in terms of memory span, the number of sequences correctly recalled, and total score. In Table 4.1, means (M), standard deviations (SD) and coefficient of variations (CV) are shown. The values presented in Table 4.1 highlight that the performances exhibited by the participants are different. The subjects outperformed in the *Corsi-Block* test. They indeed reached a greater memory span, recalled more sequences, and reached a higher total score. Furthermore, the values of the CVs are small and similar between the two experimental conditions. The CV was lightly higher for the total number of sequences correctly recalled (0.18 vs 0.15) and the total score (0.28 vs 0.26) in the *Audio-Corsi*, but the value was equal for the memory span (0.14 for both the *Corsi-Block* and the *Audio-Corsi*). The data followed a normal distribution and the homogeneity of variances between the conditions was also guaranteed for the span (Bartlett's statistic: 0.8, df=1, p=0.37), the total number of sequences correctly recalled (Bartlett's statistic:0.623,df=1, p=0.43) and the total score (Bartlett's statistic:3.52, df=1, p=0.06). The difference between the *Corsi-Block* and the *Audio-Corsi* was significant in terms of memory span (t=5.26, df=15, p <0.001, Cohen's d=1.41), sequences correctly recalled (t=7.03, df=15, p <0.001, Cohen's d= 1.75), and the product between the two (t=6.49, df=15, p <0.001, Cohen's d=1.41). Results of the comparisons are shown in Figure 4.3.

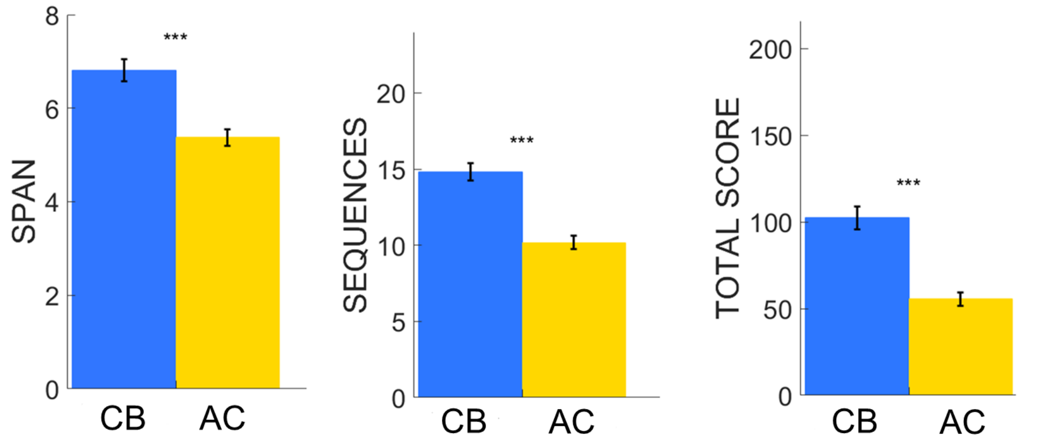


Figure 4.3 **Performances in both experimental conditions**. The panel on the left refers to the memory span, the middle panel to the number of sequences correctly recalled, the panel on the right to the total score. Three asterisks (***) indicate p <0.001. The label CB indicates the performances in the *Corsi-Block*, while AC in the *Audio-Crsi*. As shown in the figure, the subjects outperformed in the *Corsi-Block* paradigm.

Table 4.1 Performances in both experimental conditions in terms of memory span, mean number of correct sequences and the product between the two. In the Table the means, the standard deviations (SD) and the coefficient of variations (CV) are reported for both experimental conditions.

	Corsi-Block			*Audio-Corsi*		
	mean	SD	CV	mean	SD	CV
Memory Span	6.81	0.91	0.14	5.37	0.72	0.14
Number of Correct Sequences	14.81	2.26	0.15	10.19	1.83	0.18
Total Score	102.37	26.31	0.26	55.5	15.92	0.28

4.1.5 Discussion

This study presents a novel technological solution for the evaluation of audio-spatial WM abilities employing 3D spatial audio technology. Our aim here was to test and provide to researchers and clinical researchers an adaptation of the *Corsi-Block* test to be used with auditory stimuli, which, therefore, can also be employed in case of blindness. To this aim, we administered the *Corsi-Block* paradigm (Kessels et al., 2000) and the proposed *Audio-Corsi* to 16 sighted individuals, and their performances were compared between the two experimental conditions. The memory span, the number of correct sequences, and the total score were evaluated. In the *Corsi-Block* test participants reached a greater memory span and recalled more sequences (see Figure 4.3) than in the *Audio-Corsi* test. Not only this result provides an important reference for future assessments of audio-spatial memory, but it also throws light on the differences between sensory modalities in the storage and recall of spatialized items. Moreover, the system developed in the current research constitutes an effective and technologically innovative method for the evaluation of audio-spatial WM skills easily implementable and reproducible by researchers in the field of basic and clinical neuroscience.

Overall, we tested the validity of the *Audio-Corsi* test by comparing variability performance with the *Corsi-Block* test. By taking the coefficient of variation (CV) as a measure of variability in performance (see section 4.1.2), we measured the overall accuracy of each test in providing the memory span in the group of tested participants. As reported in table 4.1, the CVs for both conditions were less than 1, thus indicating that the estimate provided by both tests was quite precise. Furthermore, the CVs for the memory span were equal for both the *Audio-Corsi* and *Corsi-Block* tests, confirming that the auditory sources were memorized and recalled similarly across participants in both tests. The rationale behind this measurement relies on providing a reliable and robust task to measure audio-spatial memory. Although we observed differences between the tests, the high precision reached by the *Audio-Corsi*, comparable to the *Corsi-Block* test, strongly suggests that our methodology can be used as a valid equivalent spatial memory test. Moreover, thanks to this experiment, we have been able to identify and quantify, for the first time, an equivalent audio-spatial memory span, lower than the visual span that could be used as a reference for further studies.

As mentioned, we observed a difference in the span between the two conditions. In the *Corsi-Block* test, the participants reached, on average, a visual span between 6 and 7, in line with previous literature (Kessels et al., 2000). In the *Audio-Corsi* instead, we measured a memory span between 5 and 6 (one point less than the *Corsi-Block*). This can be ascribed to the different roles of vision and audition in the processing of spatial locations, also reflecting different mechanisms associated with audio and visual-spatial memory. Some studies already demonstrated that localization abilities are worse for sounds compared to visual objects (Zimmer and De Vega, 1996), mostly because of the different spatial resolution of the two sensory systems. While spatial processing is mainly achieved through vision, on the other end of the continuum, auditory processing is mostly related to the processing of temporal information, such as rhythm (Henneman, 1952). In this context, it has been argued that vision dominates spatial processing (King et al., 2000; Kitagawa and Ichihara, 2002; Pouget et al., 2002), not only for localization but also for memory tasks.
Conversely, audition dominates temporal processing (Guttman et al., 2005). Along these

lines, Lehnert and Zimmer demonstrated that visual stimuli are better localized than the auditory stimuli, even if presented in the same spatial layout (Lehnert and Zimmer, 2006). Furthermore, visual inputs lead to more accurate spatial representations than do auditory inputs (Paivio, 1991). However, in our task, all participants were able to localize sounds after the training. It suggests that it is not localization ability itself that influences performance in our task. In the recalling process, audio-spatial memory might be more prone to decay, given the weaker role of audition in the processing of spatial information. This result throws light on the role of the different sensory modalities in the processing of spatial information as well as in higher cognitive processes such as memory.

Beside the difference given by the encoding sensory modality, set-up properties such as the interaction with the experimenter might have influenced the differences between the two tests. In the case of the *Corsi-Block* test, the participant is asked to observe the experimenter indicating the cubes to be memorized. This procedure might induce a motor priming effect that influences performance. In detail, the observation of a finger movement might elicit an analogous motor planning response in the observer; that is, the observed actions are automatically mapped onto the internal motor repertoire of the observer (Iacoboni et al., 1999). It could lead to facilitation when the observed and executed movements are congruent (Brass et al., 2001, 2000). In the *Corsi-Block* paradigm, finger movements of the experimenter and the participant are not equal but mirrored. Thus their overall congruence could have led to a motor priming effect. This phenomenon leads to a slight reduction of the memory load, from which the subjects could have benefit in the recalling phase of the sequences in the *Corsi-Block*. In the *Audio-Corsi* test the interaction between experimenter and participant is not necessary. In this sense, the *Audio-Corsi* could be an optimal methodological solution to pursue further research investigating the role of joint action in clinical tests involving more than one agent.

Regarding the apparatus, the *Corsi-Block* and the *Audio-Corsi*, differ in the number of stimuli used: 9 blocks vs. 6 sounds. Although this aspect should not have influenced performance as most of the participants did not reach the maximum number of recalled items, their disposition changes systematically between the two conditions, especially regarding the encoding and response phase. In the *Corsi-Block*, the nine blocks were in front of the participant, during both stimuli presentation and participants' recalling. In the *Audio-Corsi*, the sounds were presented around the participant's head while their dominant index finger was kept on the reference button. It could have led the participant to refer to a head- or hand-centered reference frame, thus spatial representation could have been anchored to different body parts depending on individual participant strategy (Arbib, 1991). However, since these representations are still within the egocentric frame of references, the potential anchoring to one or the other should not have influenced performance. Moreover, the training before the *Audio-Corsi* was performed to make sure participants learned the correspondence between the spatial disposition of buttons and sounds. Further studies, perhaps investigating the contribution of egocentric and allocentric frames of reference, could be pursued with the system proposed here by focusing on the spatial relationship between sound disposition regardless of the participant's position.

The *Audio-Corsi* might have automatically required verbal coding compared to the *Corsi-Block* procedure. The participants might have recalled the sequences by verbalizing the order of the buttons or by associating each button with a number. At the end of the experimental session, only a few participants (less than 3) reported the attempt to recall the sequences by verbalizing them with the strategies described above, similar to previous observation with a haptic adaptation of the *Corsi-Block* test. Nevertheless, in our task, the time interval between the memorization and the recalling of the sequence was quite short, thus suggesting that in case the verbal strategy were to be adopted, participants would be unlikely to code the items in the series verbally.

Overall, the *Audio-Corsi* presents several advantages compared to the *Corsi-Block* test. From a methodological perspective, the *Audio-Corsi* allows the experimenter to control the Inter-Stimulus Interval (ISI) and the presentation time for each stimulus. Contrary, in the *Corsi-Block* test, there is no fixed timing since the procedure is not automatic but depends on the experimenter's ability. A solution to this issue has been proposed by developing a tablet-like version of the *Corsi-Block* test (Brunetti et al., 2014). The authors pointed out several limitations of the *Corsi-Block* test, like the possibility of the examiner to change the finger used to show the stimuli and the pace at which the tapping is performed. Along these lines, our *Audio-Corsi* provides a solution to these issues in the context of audio-spatial memory. The possibility to manipulate temporal stimulus properties provides the researcher with the means to investigate the role of presentation time in spatial memory, such as the duration of the stimulus or the ISI.

To conclude, in the current research, we compared visual- and audio-spatial memory abilities in sighted individuals employing the *Corsi-Block* paradigm and our auditory adaptation, namely *Audio-Corsi*, based on VR simulations. The performance was higher in the *Corsi-Block* test, in terms of memory span, sequences correctly recalled, and the total score (see Figure 4.3). The low variability observed in both tasks (see table 4.1), strongly supports the proposed VR apparatus as an efficient methodology to measure audio-spatial memory span. The system presented in this research does not only shed light on audio-spatial memory abilities but may also constitute the starting point for the design of novel clinical procedures, addressed to evaluate high cognitive skills in blind and visually impaired individuals. Being the exact adaptation of the *Corsi-Block* test for the auditory modality, the *Audio-Corsi* test, may be used for the early diagnose of cognitive impairments or delays in early visual deprived children, to foster the improvement of the lacking skills and, consequently, their quality of life and social inclusion.

4.2 *Audio-Corsi* test and memory processes: case study with blind and sighted adults

After having shown the efficiency of the auditory VR system for the evaluation of audio-spatial memory abilities, in this study, we compared audio-spatial memory skills in blind and sighted individuals with the same apparatus. Technologies based on virtual reality simulations represent indeed useful tools to investigate high-level cognitive functions (Picinali et al., 2014), and many validation works highlighted the superiority of virtual sounds for guidance, especially in case of early visual deprivation (Loomis et al., 2002). Indeed, interactive VR navigation helps blind individuals in building accurate representations of the surrounding environment. Within the WM domain, the *Corsi-Block* task has been useful to shed light on theoretical assumptions of the model proposed by Baddeley (Baddeley et al., 1986). This paradigm serves to evaluate serial-spatial memory because it requires the recall of both the positions of the blocks in a sequence and their order of presentation (Rudkin et al., 2007). The involvement of the central executive component is surely higher when the task requires the recalling of the sequences in reverse because more strategical control is needed (Vandierendonck et al., 2004). The cognitive processes underlying the *Corsi-Block* have been long debated (Logie and Marchetti, 1991; Rudkin et al., 2007). Some studies pointed out that the task would rely on a purely serial spatial component of visuo-spatial WM because the performances are influenced by the presentation of spatial but not visual tasks (Logie and Marchetti, 1991). Therefore, the evaluation of spatial-memory abilities in the blind can help the identification of the nature of the mechanisms underlying the paradigm. In fact, it is not well clarified the effect of early visual deprivation in spatial-memory abilities. Some studies highlighted that congenitally blind people are able to perform (sometimes better than sighted controls), tasks requiring the storage of distances and positions haptically explored (Thinus-Blanc and Gaunet, 1997; Vecchi, 1998). Conversely, other studies have shown that the blind show specific limitations to simultaneously treat and maintain different spatial patterns, supporting the dominance of visual perception in this context (Vecchi et al., 2004). Specifically, for tasks involving both the maintenance (passive) and processing (active) of spatial information, blind participants perform worse than controls (Cattaneo et al., 2008). In this study, we addressed these issues by studying the effect of early visual deprivation on the *Audio-Corsi* by comparing blind and sighted individuals' performances. Two conditions were introduced based on the sounds employed: semantic (i.e., meaningful sounds) and non-semantic (i.e., pink noise). Participants were requested to recall the sequences in the same order of presentation or in reverse, to elucidate the degree of involvement of executive resources (i.e., cognitive load) when the sequences have to be memorized in reverse. More precisely, in the current study, we wanted to answer the following research questions:

1. To which extent does the cognitive load influence audio-spatial memory abilities in an acoustic virtual environment?

2. What is the role played by visual experience on the task?

Blind individuals show limitations when asked to manipulate spatial information stored in memory (Cattaneo and Vecchi, 2011; Setti et al., 2018; Vecchi et al., 2005). Nevertheless, this topic has never been investigated in a system based on an acoustic VR. In line with Ruggiero and colleagues (Ruggiero et al., 2009), our hypothesis was that the absence of visual experience could affect audio-spatial memory abilities, especially in backward order. The request to combine the sounds in compact representations and the need to use visuo-spatial imagery strategies as the length of the sequences increases should affect the performances of the blind. Unlike the previous studies, here we employed an acoustic VR to study these issues, where the sounds were not in front of the participants but displaced around their heads.

4.3 Task and Procedure

Sample

The study involved 9 sighted (5 females, mean age ± SD: 30.11 ±6.89 years old) and 9 congenitally blind individuals (7 females, mean age ± SD 31.11 ±8.08 years old). Both groups reported no cognitive impairment. Blind adults were recruited from our institute database, and the clinical details of the blind are shown in Table 4.2. The experiments were approved by the ethics committee of the local health service (Comitato Etico, ASL 3, Genoa, Italy) and performed in accordance with the declaration of Helsinki. Adult informed written

Table 4.2 Participants' clinical details

PARTICIPANT	AGE	GENDER	RESIDUAL VISION	PATHOLOGY
S1	20	F	No Vision	Retinopathy
S2	32	F	No Vision	Retinopathy
S3	29	F	No Vision	Leber's Amaurosis
S4	26	F	No Vision	Retinopathy
S5	27	M	No Vision	Retinopathy
S6	46	M	No Vision	Cataract
S7	30	F	No Vision	Retinopathy
S8	28	F	No Vision	Glaucoma
S9	42	F	No Vision	Glaucoma

consent for the study was obtained in all cases. The experiment did not have a fixed duration, since it depended on the ability of the participant. However, it lasted 25 minutes on average.

Experimental Procedure

The test was carried out with the acoustic VR system described in section 2.2. We adapted the *Corsi-Block* test (Corsi, 1973) to the auditory domain. In the test, participants were asked to recall sequences of sounds of increasing length, starting from 2 up to 9 sounds. The paradigm was divided into 2 experimental conditions: non-semantic and semantic. In the non-semantic condition, the auditory stimulus was pink-noise (time length: 2.5 seconds) for all the locations. In the semantic condition, we used meaningful sounds (time length: 2.5 seconds), as shown in Figure 4.4. The sounds were presented binaurally, and the RMS level was calibrated to be the same across the various signals. The interaural intensity, for the different locations, was equal to the one presented in the previous study. The right and left positions, closer to participants' head, were simulated by an interaural intensity difference of 20 dB (about 55 and 75 dB sound pressure level for each ear). The north-east and north-west locations were simulated by an interaural intensity difference of 10 dB (about 34 and 44 dB sound pressure level for each ear). The stimulus in the frontal position was presented binaurally at an equal intensity (about 80 dB sound pressure level). Regards the back location, there was an intensity difference of 8 dB (about 32 and 40 dB sound pressure level for each ear). The two conditions were counterbalanced across participants. Half of the subjects were first tested with the non-semantic condition, while the other half started with the semantic condition. None of the participants had seen the apparatus before the beginning of the test, and the group of sighted entered the experimental room blindfolded.

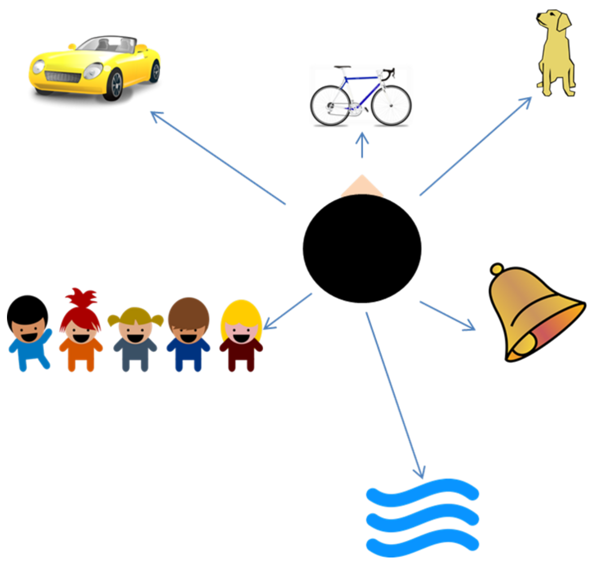


Figure 4.4 **Sounds' disposition in the semantic condition.** The picture shows the sounds used in the semantic condition. In the simulation, the subject is at the center of the virtual environment and listens to spatialized sounds as coming from 6 locations: north-west, north, north-east, east, south, and west. The sounds placed in these positions were, respectively: the car, the bike, the dog, the bells, the water, and the kids while playing. The images used to create this picture were downloaded by a royalty-free images web archive (https://publicdomainvectors.org/)

Once sat, each subject touched the keyboard, trying to become confident with its size and buttons' displacements. Then, the experimenter and the participant counted the red buttons together (see Figure 2.10). They were instructed to consider the central blue button as spatial reference (see Figure 2.10). The test started with a training to learn the virtual environment. In the non-semantic condition, the training helped the participants to understand how the sounds are perceived based on their spatial locations. In the semantic condition, instead, it also served as a way to understand the virtual environment. In the 3-minute training, the participants freely touched the red buttons, and after each touch, a sound was reproduced through headphones. Afterward, the experimenter checked whether the participants had understood the environment. The sounds were emitted, one by one randomly, from the 6 locations. The participants pressed the buttons corresponding to the positions from which the sounds appeared to come. Then, the test started. In the semantic condition, before the training, the subjects listened to the sounds one by one and identified them. In the test phase, the participants listened to sequences of sounds of increasing length, 3 sequences per length. If at least one sequence was correctly recalled, the test went on; otherwise, it stopped. Based on the recalling modality, we defined two blocks per each condition: forward and backward.

In the first, the sequences were recalled in the same order of presentation; in the second instead, in reverse. After each sequence, the participant placed the dominant hand on the reference button. The experimenter did not communicate whether the subject remembered well the sequence. Only the change in sequence length was verbally communicated. The participants wore a pair of Sennheiser HD-650 headphones to deliver the sound stimuli. To see whether the training was effective for the learning of the virtual environment, in both experimental conditions, we compared the percentage of single sounds recalled by the groups in both conditions.

4.3.1 Data Analysis

To quantify the subjects' performance, we evaluated

1. The memory span;
2. The number of sequences correctly recalled;

The span was calculated according to the formula 3.1.

4.3.2 Statistical analyses

Statistical analyses were carried out in Matlab (Matlab R2017, The Mathworks) and R (RStudio 4.3.2 version). The normality of the data was checked with the Jacque-Brera test (*jbtest()* Matlab function). The analysis of the percentage of sounds correctly recalled after the training was performed in R. We designed a mixed measures ANOVA based on permutations (*aovp()* function), with *Group*, either blind or sighted as between factor and *Condition*, either semantic and non-semantic as within factor. Performance analyses were carried out separately for the parameters listed above. We designed, for each of them, a 3-ways 2x2x2 mixed measures ANOVA with *Group* as between and *Condition* and *Process* (either forward or backward) as within factors. Post-Hoc analyses were carried out with two-tailed Student's t-tests both paired and unpaired (*ttest()* and *ttest2()* Matlab functions). *Bonferroni* correction was used to correct for multiple comparisons, and $p<0.05$ was considered significant. Cohen's coefficient was evaluated to quantify the effect size (Cohen's d).

4.3.3 Results

Results related to the single source localization are shown in Figure 4.5. Analysis of

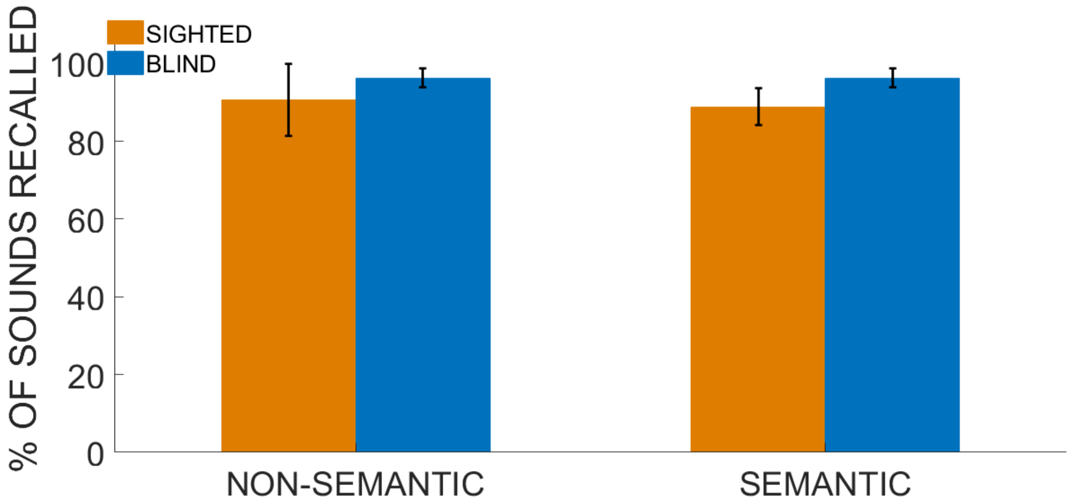


Figure 4.5 **Percentage of sounds recalled after the training.** Data are presented as mean and standard error. The figure shows the percentage of the sounds recalled after the training, by both groups in both experimental condition. we did not find a significant difference between the groups in the number of sounds recalled.

variance did not reveal any significant effect (*Group*: p=0.56 n=90, Cohen's d=0.12 *Condition*: p=0.48 n=83, Cohen's d =0.2 *Group*Condition*: p=0.76 n=70, Cohen's d=0.3). This suggests that, on average, all the groups recalled the same number of auditory sources. Results for the memory span are shown in Figure 4.6. Analysis of variance only revealed a significant main effect of the *Condition* (F(1,16)= 12.96, p <0.001, Cohen's d=1.23), of *Group* (F(1,16)=12.96, p <0.01, Cohen's d=2.5) and interaction *Group*Process* (F(1,16)=5.1, p=0.0382, Cohen's d=0.73). We did not find any significant interaction *Group* Condition* (F(1,16)=1.36, p=0.24, Cohen's d=0.29), *Condition*Process* (F(1,16)=1.78, p=0.31, Cohen's d=0.33), *Group*Condition*Process* (F(1,16)=0.09, p=0.77, Cohen's d=0.003) nor main effect of *Process* (F(1,16)=0.61, p=0.45, Cohen's d=0.007). Post-Hoc analyses, showed that the span reached by the subjects was higher in the semantic rather than non-semantic condition, apart from the group and the process (two-tailed unpaired Student's t-test: t-value=2.034, p=0.044, df=70, Cohen's d=0.48). Moreover, sighted participants memorized longer sequences, compared to the other group, regardless of the recalling request (two-tailed unpaired Student's t-tests: *Forward*: t-value= 3.48, p<0.01, df=34 Cohen's d=1.12. *Backward*: t-value=1.03, p<0.001, df=17, Cohen's d=1.74). Furthermore, only the blind showed a significantly higher span in the forward rather than backward process (two-tailed paired Student's t-test: t-value=2.99, p=0.033, df=17, Cohen's d=0.7). Analysis of variance for the number of sequences correctly recalled revealed the same pattern of results. We

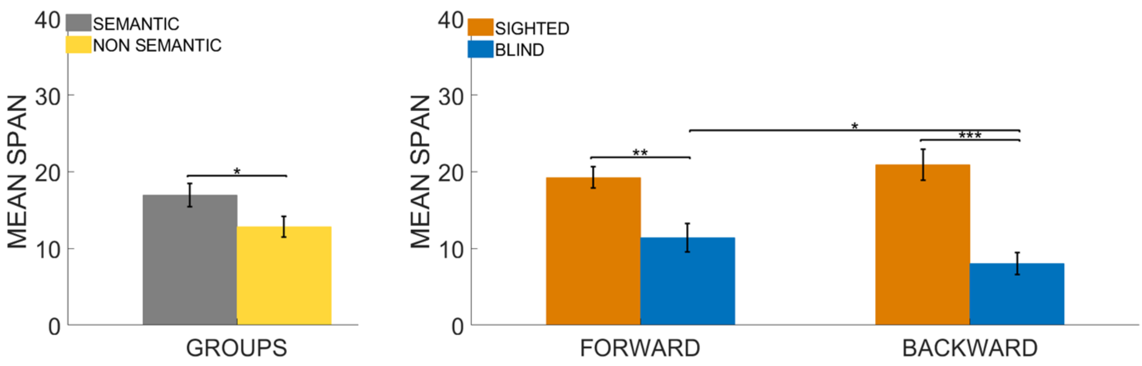


Figure 4.6 **Memory Span.** Data are presented as the mean and standard error. Results point out that the span was higher in the semantic condition (panel on the left). The sighted outperformed apart from the experimental condition, and only in the group of the blind, we found a significant difference in the span between the forward and backward processes (panel on the right). The asterisks indicate statistical significance (one asterisk (*) represents p<0.05, two asterisks (**) represent p<001 and three asterisks (***) represent p<0.001).

only obtained a main effect of the *Condition* (F(1,16)=33.49, p<0.001, Cohen's d=1.42), of the *Group* (F(1,16)=14, p<0.01, Cohen's d=3), and a significant interaction *Group***Process* (F(1,16)=5.91, p=0.023, Cohen's d=0.48). We did not find any significant main effect of the *Process* (F(1,16)=2.01, p=0.17, Cohen's d=0.27) nor significant interaction *Group***Condition* (F(1,16)=0.35, p=0.56, Cohen's d=0.1), *Condition***Process* (F(1,16)=0.9, p=0.4, Cohen's d=0.18) nor *Group***Condition** *Process* (F(1,16)=0.23, p=0.64, Cohen's d=0.19). Post-Hoc analysis firstly revealed that participants recalled more sequences in the semantic condition (two-tailed unpaired Student's t-test: t-value=2.33, p=0.022, df=70, Cohen's d= 0.549). Sighted participants memorized more sequences than the other group (unpaired two-tailed Student's t-tests:*Forward* t-value=3.48, p<0.01, df=34, Cohen's d=1.15, *Backward* t-value=5.43, p<0.001, df=34, Cohen's d=1.8) and only the blind recalled more sequences in the forward rather than backward order of presentation (unpaired Student's t-test: t-value=3.08, p=0.027, df= 17 Cohen's d=0.73).

4.3.4 Discussion

In the current work, we compared audio-spatial WM abilities in sighted and congenitally blind individuals. Both groups were tested with the *Audio-Corsi* divided into two experimental conditions: semantic and non-semantic. The results show that regardless of the group and the order of recalling, participants outperformed in the semantic condition. Furthermore, irrespective of the experimental condition, the group of sighted performed better in both

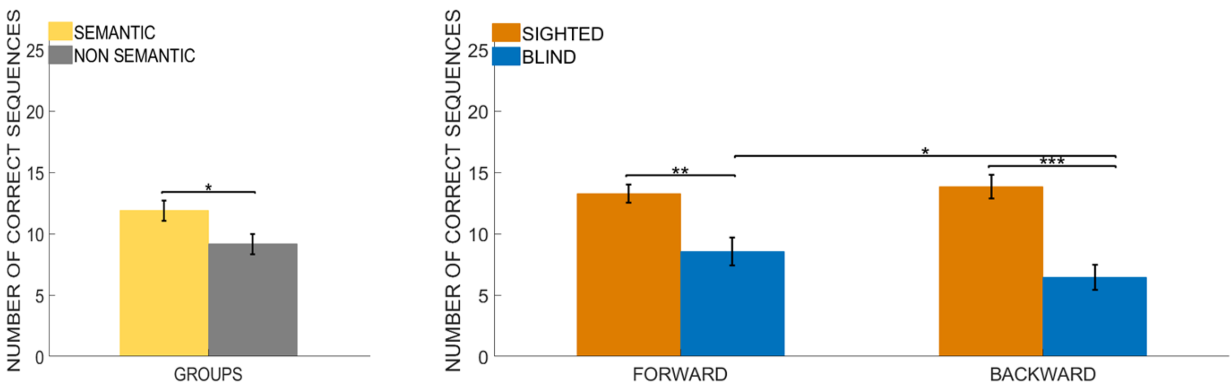


Figure 4.7 **Sequences correctly recalled.** Data are presented as the mean and standard error. Results point out that, in the semantic condition, the participants recalled more sequences (panel on the left). The panel on the right highlights that the sighted outperformed concerning the other group. Additionally, only the group of blind shows a significant difference in the number of sequences memorized between the forward and backward processes. The asterisks indicate statistical significance (one asterisk (*) represents p<0.05, two asterisks (**) represent p<001 and three asterisks (***) represent p<0.001).

orders of recalling, and only the group of blind exhibited a significant difference between the forward and backward orders of presentation (see Figures 4.6 and 4.7). The semantic condition was easier to be performed. Once the locations of the sounds were discovered, the participants did not need to pay attention to sounds' location; they could press the button which corresponded to that specific sound or use verbal codes. Conversely, in the non-semantic condition, the association between the button and the sound was less obvious. The pattern of results shown in the right panels of Figures 4.6 and 4.7, is likely due to a different use of the spatial relations among the sounds and not to a different ability to localize them (see Figure 4.5). The *Corsi-Block* test indeed acts on the VSSP component of the WM system and therefore requires the generation and the processing of mental images (Taevs et al., 2010). As stated by Vandierendonck and colleagues (Vandierendonck et al., 2004), in the *Corsi-Block*, upon presentation of a series of block positions, a representation of the path is constructed and maintained in visuo-spatial working memory. In the *Audio-Corsi*, participants were also asked to learn sequences of spatialized sounds of increasing length. It is, therefore, likely that subjects were inclined to use a spatial strategy to accomplish the task. Wanet-Defalaque (Vanlierde and Wanet-Defalque, 2004) underlined the importance of the strategy used to accomplish a certain task, especially in the case of congenitally blind people. Cornoldi et al. (Cornoldi et al., 2009) deeply investigate this topic by asking sighted and blind individuals to explore 2D matrices (5x5), by following specific instructions to recall

the whole pathway or just the final point. The authors concluded that when the task is simple or quite suitable for being verbally mediated (as in the request to recall only the last position of the pathway), blind individuals can reach a good level of performance. Nevertheless, when the task is more demanding, requiring the active manipulation of spatial information or the use of mental spatial representation, the performances exhibited by the blind drop. In the *Audio-Corsi*, the use of a pure verbal strategy was more difficult as the sequences increased in length, and both groups had to rely on the mental representation of the pathway (i.e., the positions of the sounds composing the sequences). Previous works highlighted that vision allows the coding of spatial information in the form of global, externally based representations. At the same time, blind people encode spatial information in the form of "route-like" representations (Thinus-Blanc and Gaunet, 1997). This is a direct consequence of the nature of haptic and auditory exploration, based on sequential processing of space (Cornoldi et al., 1979; Cornoldi and Vecchi, 2003). Thus, although temporal processes are important in the *Corsi-Block* task (i.e., the sequential encoding of block positions), the results, in line with Ruggiero and co-workers (Ruggiero and Iachini, 2010), suggest that what is crucial in this paradigm is the spatial component. As the length of the sequences increased, sighted participants performed better than blind individuals, namely, they combined the spatial information into unified and functional representations (see Figure 4.6 and 4.7).
In the *Audio-Corsi* test, participants recalled sequences of spatialized sounds of increasing length. As the number of sounds to be memorized increased, the performance of the blind worsened, especially in the backward order (see Figures 4.6 and 4.7 for details, right panels). The request to recall the sequences in reverse implied a higher cognitive load. Thus, the amount of involvement of executive resources might have also had an indirect effect on blind's performances. Some authors argued that there should be no difference in sequence recalls between the forward and backward order of presentations. In both cases, the participants are just asked to recall a sequence mentally represented like a configuration. However, previous research pointed out that the backward span is lower than the forward span (Iachini et al., 2005; Rudkin et al., 2007; Vandierendonck et al., 2004). Moreover, the request to reverse the sequence is strictly correlated with active spatial paradigms mental rotation and spatial inference (Ruggiero et al., 2008).
In conclusion, in the current research, we compared for the first time, through an acoustic VR simulation, audio-spatial memory abilities in blind and sighted individuals. The results highlight that the lack of visual experience could limit the strategies available to process sound sequences, particularly in the demanding backward order. Congenitally blind participants might suffer more from the cognitive load imposed by the task and from the indirect need to

use the spatial relations among the sounds composing a sequence.
From a technological point of view, we observed that visually impaired subjects easily interacted with the acoustic VR system. They were able to localize the sounds after the training and to recall the sequences in the test phase. Therefore, this apparatus may constitute the starting point for the design of experimental and clinical procedures for the evaluation, and eventually rehabilitation, of high-level cognitive abilities (e.g., spatial memory) in the blind population.

Chapter 5

General Discussion

During my three years of PhD, I worked on the development of acoustic devices and experimental paradigms to be used in case of early visual deprivation. The main goal was to use these technologies to design auditory methods for the assessment of WM and spatial functions, in sighted and congenitally blind individuals of all ages. Although blindness affects millions of people, there is a lack of devices that use sensory modalities other than vision to study, clinically assess, and eventually train, high-level cognitive skills such as spatial memory. When studying cognitive functions involving auditory spatial information, multiple auditory spatial locations are required, and thus multiple speakers or simulation of spatialized sounds via headphones. The solutions here developed allow the measurement of various aspects of acoustic spatial perception and auditory working memory. My main scientific goal was to investigate how these abilities develop in the first period of life and the impact of early visual deprivation. These devices indeed were thought to be customized and suitable for people of all ages, both sighted and visually impaired. Validated clinical procedures and games were adapted to the auditory modality (i.e., the *Corsi-Block* test and the card game *Memory*) to test the effectiveness of these devices for the evaluation and eventually rehabilitation of cognitive impairments. From a technological point of view, we developed novel devices that allow the spatial emission of auditory stimuli, described in the next paragraph.

5.0.1 Technological solutions developed during the PhD

The first device used in my PhD is *ARENA2D*. The latter is an acoustic tablet (50 x 50 x 10 cm), composed of 25 haptic blocks, each covered by 16 tactile sensors (4x4 matrix) and with a speaker in the center (see Figure 2.1 in section 2.1). This device allows the serial

emission of spatialized sounds and the simultaneous recording of the touches on its surface. The acoustic stimuli can be played independently, based on the software or programming language used, or when the user touches a specific block; in this case, the sound might be emitted from the selected block or from another one.
This device has already been employed to evaluate the accuracy in auditory localization tasks (Cappagli et al., 2017a). In the tests presented in this thesis, besides the identification of sounds' origin, participants were asked to keep in mind stimuli' locations for immediate or later recalling. In the tasks where *ARENA2D* was employed, the participants were requested to explore the device while learning the spatial positions occupied by the stimuli. This system has been used to investigate how blind and sighted individuals explore, learn, memorize, and manipulate a complex acoustic spatial layout. The resulting representations could be static (i.e., the sounds occupied always the same position) or dynamic (i.e., the spatial layout changed during the experiment because new sounds were added or removed in turn). *ARENA2D* is therefore particularly suitable when the request is to know how a complex audio-structure is explored and memorized by the users. Finally, the device permits both the passive listening of either single or multiple sounds or the active interaction between the user and the technology (i.e., when the users touch the surface while listening to the stimuli). To summarize, *ARENA2D* allows:

1. A good spatialization of sounds, mostly thanks to its size and the number of blocks;
2. Active and passive interaction with the technology;
3. The possibility to use up to 25 sounds of different nature (i.e., non-semantic vs. semantic);

We used this technological solution to design an acoustic version of the card game *Memory* and an audio-spatial task for blind and sighted individuals taking inspiration from the *Corsi-Block* paradigm. In the *Audio-Memory*, participants were asked to associate stimuli displaced on the *ARENA2D* surface. As in the original game, the subjects were required to find all the matches among a set of sounds. This test, however, is not only an adaptation of a game, which nevertheless makes it suitable for children. Thanks to the *Audio-Memory*, we have been able to evaluate how the participants explored the spatial layout built with the sounds, their mnestic and audio-spatial skills, and, in the case of the children, the role played by the WM development (i.e., the phonological loop). With *ARENA2D* we also created a meaningful spatial layout (see Figure 3.10, section 3.3) that the participants (both congenitally blind and sighted) learned and memorized. They were then tested with sequences of spatialized sounds of increasing length with the same sounds that composed the acoustic scene. This experiment aimed to investigate how the relations among the stimuli are used for sequences' recalling, especially after it was discovered that the sounds belonged to a coherent scene. *ARENA2D*, however, suffer from some limitations:

1. It is not portable;
2. It does not allow the parallel emission of sounds (i.e., more stimuli at the same time);
3. It is costly;
4. In noisy environments, it could be hard to localize the sounds precisely;

These points highlight that it may be difficult to use this tablet in clinical settings or schools and places attended by children. We therefore designed a system based on an acoustic VR (see Figure 2.10 and 2.8). In sections 2.2 and 2.2.4 the technical details are explained. This system is cheaper than *ARENA2D*, more portable, and the sounds are played through headphones, thus preventing more the influence of the surrounding noise. The apparatus consists of a set of virtual sounds (6 based on pilot studies displaced around the user's head, see Figure 2.8 in section 2.2.2) and a custom-made wooden keyboard provided of 6 buttons replacing sounds' locations and a central button as a spatial reference. The use of this technology for exploring and evaluating spatial hearing processes and high level cognitive mechanisms can be considered as being particularly novel.

5.1 Development of audio-spatial memory abilities

ARENA2D was employed to study how audio-spatial WM skills develop during the first period of life and to which extent the changes in the phonological loop functionality influence them. This topic was investigated with the experimental paradigm described in section 3.1. Previous research pointed towards a developmental shift from phonological to semantic processing of words that occurs around the 8th year of life (Dewhurst and Robinson, 2004). Six and seven year-olds, are perhaps prone to encoding the animal's name phonologically (or acoustically), rather than semantically. Animal call and spoken names have different acoustic features, so that stimuli coupling that is driven by memorization of locations may be more difficult for this group of children. Children who are 10-11 years-old instead encode a word focusing on its meaning and the concepts behind (Hitch et al., 1989). After passing the age of 7, the rate of rehearsal increases, enabling children to retain a greater amount of verbal material (Gathercole et al., 2004). Before that age, the rehearsal of auditory speech material does not spontaneously occur (Gathercole et al., 1994, 1999). Thus, thanks to the better use of the rehearsal, it is likely that older children could remember the locations of the spoken words better than the younger peers, likely by preventing decay of verbal material through rehearsal (Hulme and Tordoff, 1989). When the sounds to pair are identical (two animal calls) 6-7 year-olds can match them through purely acoustic properties.

More generally, the development of the phonological loop influences visual item encoding. Six and seven year-olds only rely on the visuo-spatial sketchpad to recall visual stimuli, focusing on physical features. Older children instead usually recall visual inputs into a phonological form by employing the rehearsal process (Pickering, 2001). In other words, they memorize the objects by giving labels to them; for example, a bottle is remembered because they name the object as "bottle" and not because of its physical properties as a cylindrical shape. According to our results, these processes are perhaps valid in the auditory domain. The focus on the acoustic features of a sound refers to the physical characteristics of the stimulus. Older children pay more attention to the concept behind the sound, so they may have labeled the stimuli based on their meaning (i.e., the "bellow" corresponds to the cow). We thus speculate that the results, suggest that after the 8th year of life, children rely on using phonological loops to label sounds. In this sense, the same processes could be valid for both the visual and auditory systems. Further research is needed to specifically evaluate the use of the labeling objects for audio-spatial memory, for instance, by testing whether older children are more prone to using this memory strategy compared to their younger peers and whether this extends to different categories of objects. However, the developmental stage likely influenced how the spatialized sounds were stored and matched.

By testing children with the *Audio-Memory*, we also investigated the influence of the developmental stage on the adopted exploration strategy to unveil how children build spatial representations in the auditory domain while performing tasks. Previous research (Blades and Spencer, 1994) pointed out that children start integrating metrical and categorical representations of space at 4 years of age (Newcombe, 2002). Moreover, before the age of 7 years-old, they are not able to properly integrate spatial metrics, at least for the visual modality (Vasilyeva and Lourenco, 2012). The current work highlights similar difficulties when children explore complex audio-spatial structures.The greater use of an audio-anchor indicates that 6-7 year-olds do not account for the mutual relations between the elements; for example, the dog is next to the rooster and under the sheep. As already discussed in the section 3.1.4, younger participants do not properly integrate the spatial positions in relation to nearby items. There is significant evidence for unitary storage of spatial information independent of the encoding sensory modality (Lehnert and Zimmer, 2006). Neurophysiological evidence confirmed this view by revealing that the generation of spatial representations involves supramodal brain areas devoted to spatial processing (Röder et al., 2001). The results here presented suggest that the deficits young children show in using the respective positions of multiple landmarks are valid for both visual and auditory domains.

5.2 Audio-spatial memory abilities in blind individuals

The technologies presented in the current thesis were employed in the comparison of audio-spatial WM abilities in congenitally blind and sighted individuals. We investigated how blind people manipulate acoustic spatial information, learn and process complex auditory structures and recall sequences of spatialized sounds.
In this section, I will compare the findings obtained with the studies on the effect of blindness on audio-spatial memory. I will refer to the studies as:

1. Audio-spatial memory task (conducted with ARENA2D and described in section 3.3)
2. The *Audio-Memory* (described in section 3.1)
3. The *Audio-Corsi* (described in section 4.2)

We first observed that audio-spatial memory skills are not strictly affected by early visual deprivation when the task demand is only the memorization of spatial information (i.e., passive memory test), and there are few items to be stored and recalled. In the audio-spatial memory task, both sighted and congenitally blind individuals recalled the same number of sequences before the spatial exploration, regardless of the nature of the sounds (i.e., semantic vs. non-semantic). Previous works that investigated spatial memory for haptic layouts in blind individuals, have drawn the same conclusions (see section 1.3). Vecchi and colleagues (Vecchi et al., 1995) tested the hypothesis that blindess-related difficulties strongly depend on the difference between passive and active memory processes. They asked sighted and blind individuals to explore wooden matrices covered by sandpaper targets. Afterward, the participants were presented with a blank matrix and asked to indicate the positions of the targets. The results show that the performances between groups did not differ.
This finding indicates that audio-spatial memory skills are not strictly affected by early visual deprivation when the task demand is the simple memorization of spatial information (i.e., passive memory test) and there are few items to be stored and recalled. In the audio-spatial task, we also looked at the influence of the semantic information on spatial memory

performances. Only after the spatial exploration the sighted exhibited better performance (see figure 3.11 in section 3.3). According to Mandler and Ritchey (Mandler and Ritchey, 1977), visual stimuli are better stored in memory and influence the pattern of results when they are part of a scene coherently organized. Semantic sounds were also used in the *Audio-Corsi*, but in that case, the performance was higher in the semantic compared to non-semantic condition (see Figure 4.6 and 4.7 in section 4.2). In the *Audio-Corsi* we focused on the investigation of audio-spatial memory to compare performances. By following the previous finding with the *ARENA2D*, we trained participants to understand that the stimuli belonged to a coherent spatial layout before the beginning of the test.

By looking at how blind individuals explore a complex acoustic layout, we observed that the resulting spatial representation was as accurate as of the one of the sighted group. In both the *Audio-Corsi* and the audio-spatial memory task, the group of blind was able to remember and recall sounds' positions at the end of the experimental session. In line with the research in the haptic modality, blind individuals can construct mental representations even in absence of an external visual input (Cattaneo and Vecchi, 2011; Cattaneo et al., 2008; Cornoldi and Vecchi, 2004; Vecchi, 1998; Vecchi et al., 2005).

The difference in the performances in the studies involving blind individuals can be ascribed to the influence of early visual deprivation on the processing of spatial information. In the audio-spatial task, in the *Audio-Memory* and in the *Audio-Corsi* we obtained a similar pattern of results: blind individuals performed poorly compared to the sighted. The organization of single items into a structured pattern is more easily accomplished when the stimuli are visually encoded rather than haptically or auditory (Davis et al., 2001). Blindness indeed specifically influences the ability to maintain and combine multiple stimuli in memory, due to the sequential nature of the perceptual experience when vision is absent since birth (Cattaneo et al., 2008; Cornoldi and Vecchi, 2003; Ruggiero and Iachini, 2010). The tendency to use an audio-anchor when accomplishing the *Audio-Memory* (see Figure 3.5) supports this view; the blind needed to refer all the locations to a fixed position to explore the spatial layout. In the *Audio-Corsi* and the audio-spatial task, the items had to be combined together into a single sequence and then recalled. The *Corsi-Block* test, upon presentation of a series of block positions, requires the construction of a representation of the path which is maintained in visuo-spatial WM for immediate recalling (Vandierendonck et al., 2004). The same is also valid for the two tasks, which required the processing of spatialized series of sounds. Visual imagery strategies may help re-coding a sequence as a structured pattern, and this might have directly affected the performances of the blind. In the audio-spatial task, the sighted participants took advantage of the spatial configuration learned during the exploration by combining them into a functional representation. Altogether, these findings indicate that the absence of visual experience confines WM abilities to sequential and slower processing of spatial information.

The cognitive load imposed by these tests also had a direct effect on the pattern of results. In the *Audio-Memory*, the participants continuously updated the spatial representation of sounds' dispositions, especially after the discovering and pairing of the animal calls. In the *Audio-Corsi* test, they memorized sequences of spatialized sounds of increasing length. As the number of sounds to be recalled increased, the performance of the blind worsened. Furthermore, in the *Audio-Corsi* the influence of the cognitive load was also more evident when the sequences had to be recalled in reverse (see Figures 4.6 and 4.7 for details, right panels). As a sequence to be recalled in backward order is presented, a reversed string is constructed and later maintained and rehearsed in WM (Vandierendonck et al., 2004). The difficulties in reversing the sequences by the blind seem to be valid in both touch and hearing, as already observed by Ruggiero and colleagues (Ruggiero and Iachini, 2010). In this type of test, indeed, there is a direct involvement of the executive processes to reverse the incoming spatial information.

Finally, the results of the studies also suggest a different use of the mutual relations of the sounds, that is, of the frames of reference, between sighted and blind individuals. The two main frames of reference used to represent the item locations in the surrounding environment are egocentric (i.e., centered to the observer's body) and allocentric (i.e., object-to-object) (Ruggiero et al., 2012). Early visual deprivation results in significant impairments in those tasks that require an allocentric representation of space (Arnold et al., 2013; Gori et al., 2013; Ruggiero et al., 2012). Thus, when the sighted participants were asked to recall the sequences of sounds, they may have used the mutual relations among them composing each sequence, especially in case of meaningful sounds, as discussed in section 2.1. The blind instead do not spontaneously use this strategy, as already seen by the tendency to start consecutive attempts by touching the same speaker in the *Audio-Memory* (i.e., the audio-anchor, see Figure 3.5). The use of the audio-anchor indeed indicates that the blind do not construct the spatial representation by using the nearby items.

To summarize, the research presented in this thesis highlight that early visual deprivation affects the active processing of the spatial information stored in memory. Previous research demonstrated these results in the haptic modality but, thanks to the technologies developed for these studies, we observed that the same conclusions could be drawn for audition. In addition to the difficulties in using the spatial relations among the sounds, with each of these tasks, we could observe different processes parallel to the manipulation of spatial information:

1. With the audio-spatial experiment ran with *ARENA2D*, we observed that blind participants can create an accurate representation of an acoustic layout and that the semantic information influences the memory performances only when the sounds belong to a coherent scene;

2. With the *Audio-Memory* we could investigate how blind individuals construct a spatial representation of a complex acoustic structure (i.e., the exploration strategy);

3. With the *Audio-Corsi* we studied, for the first time, audio-spatial memory skills in an acoustic VR. We have been able to investigate the cognitive load imposed when the experimental request was to recall the sequences in reverse, and we elucidated the role of spatial strategies in the accomplishment of this task;

5.3 The emergence of new rehabilitative technologies

Visual impairments cause spatial and social deficits at the cognitive level. Therefore, the technologies developed in this context should address one main issue: the clinical evaluation and, eventually, rehabilitation of impaired cognitive abilities. In the case of congenital blindness, these technologies should intervene in particular in the first period of life to limit, to the extent possible, the spatial and cognitive impairments due to early visual deprivation. However, most of the devices to date available are substitution devices; that is, they aim at substituting the visual information with another modality. As already discussed in section 1, these devices suffer several limitations. Starting from the limits of the existing technology, during my PhD, I developed the software of existing devices (*ARENA2D* and Audiobrush) and designed an auditory VR system (see Figure 2.9b) to evaluate e.g., spatial memory in visually impaired people of all ages. These solutions convey information in the auditory domain by means of spatialized sounds and have been tested on sighted and congenitally blind people of all ages.

5.4 Final remarks and future goals

In my PhD, I worked on the software and hardware development of three devices for the evaluation of audio-spatial memory skills in congenitally blind and sighted people of all ages.

I designed experimental procedures by adapting to the auditory domain the *Corsi-Block* test and the card game *Memory*. Our results point towards a developmental trend in audio-spatial memory skills and indicate that the same limitations encountered by blind individuals in the haptic modality are valid for audition. However, due to the small number of participants in some of the experiments here presented, further research is needed to confirm the results of this thesis. The results of the studies here presented allow the following conclusions:

1. Audio-spatial memory abilities, in typical children, are strongly influenced by cognitive development during the first period of life. The greater use of verbal coding strategies for immediate memorization and the development of spatial skills affect how a complex acoustic spatial layout is learned and explored;

2. Early visual deprivation affects the processing of spatial information, also in the auditory domain. Congenitally blind individuals can memorize sounds' positions belonging to a coherent scene, but they show limits in the manipulation of these spatial representations. Audio-spatial memory skills are similar between sighted and blind individuals until the cognitive load imposed by the task is not high. Otherwise, the performances exhibited by the blind drop due to the limits imposed by early visual deprivation in processing spatial information.

The technologies here presented might represent innovative and powerful tools for the evaluation of cognitive skills, especially when audition is one of the remaining sensory modality to cope with everyday life. These devices, as shown in the current thesis, can be easily used by children and adults with and without vision loss, thus indicating their potential inclusion in clinical and rehabilitative settings.

References

Aleman, A., Van Lee, L., Mantione, M. H., Verkoijen, I. G., and de Haan, E. H. (2001). Visual imagery without visual experience: evidence from congenitally totally blind people. *Neuroreport*, 12(11):2601–2604.

Algazi, V. R. and Duda, R. O. (2010). Headphone-based spatial sound. *IEEE Signal Processing Magazine*, 28(1):33–42.

Alloway, T. P., Gathercole, S. E., and Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child development*, 77(6):1698–1716.

Amedi, A., Raz, N., Pianka, P., Malach, R., and Zohary, E. (2003). Early ‘visual’cortex activation correlates with superior verbal memory performance in the blind. *Nature neuroscience*, 6(7):758–766.

Arbib, M. A. (1991). Interaction of multiple representations of space in the brain.

Arditi, R. and Dacorogna, B. (1988). Optimal foraging on arbitrary food distributions and the definition of habitat patches. *The American Naturalist*, 131(6):837–846.

Arnold, A. E., Burles, F., Krivoruchko, T., Liu, I., Rey, C. D., Levy, R. M., and Iaria, G. (2013). Cognitive mapping in humans and its relationship to other orientation skills. *Experimental brain research*, 224(3):359–372.

Arnott, S. R., Binns, M. A., Grady, C. L., and Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *Neuroimage*, 22(1):401–408.

Auvray, M., Hanneton, S., and O’Regan, J. K. (2007). Learning to perceive with a visuo—auditory substitution system: localisation and object recognition with ‘the voice’. *Perception*, 36(3):416–430.

Auvray, M. and Myin, E. (2009). Perception with compensatory devices: From sensory substitution to sensorimotor extension. *Cognitive Science*, 33(6):1036–1058.

Bach-y Rita, P., Hance, A., and Winters, W. (1960). Some effects of atropine and eserine on thalmo-cortical pathways in the cat. In *Proceedings of the Western Pharmacology Society*, volume 3, pages 69–79.

Bach-y Rita, P. and Kercel, S. W. (2003). Sensory substitution and the human–machine interface. *Trends in cognitive sciences*, 7(12):541–546.

Baddeley, A. (1992). Working memory. *Science*, 255(5044):556–559.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, 4(11):417–423.

Baddeley, A., Logie, R., Bressi, S., Sala, S. D., and Spinnler, H. (1986). Dementia and working memory. *The Quarterly Journal of Experimental Psychology Section A*, 38(4):603–618.

Baddeley, A. D. and Hitch, G. (1974). Working memory. In *Psychology of learning and motivation*, volume 8, pages 47–89. Elsevier.

Baddeley, A. D. and Logie, R. H. (1999). Working memory: The multiple-component model.

Beauchamp, K. G., Kahn, L. E., and Berkman, E. T. (2016). Does inhibitory control training transfer?: Behavioral and neural effects on an untrained emotion regulation task. *Social cognitive and affective neuroscience*, 11(9):1374–1382.

Beni, R. d. and Cornoldi, C. (1988). Imagery limitations in totally congenitally blind subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(4):650.

Berch, D. B., Krikorian, R., and Huha, E. M. (1998). The corsi block-tapping task: Methodological and theoretical considerations. *Brain and cognition*, 38(3):317–338.

Blades, M. and Spencer, C. (1994). The development of children's ability to use spatial representations. *Advances in child development and behavior*, 25:157–199.

Blankenship, A. B. (1938). Memory span: a review of the literature. *Psychological Bulletin*, 35(1):1.

Brass, M., Bekkering, H., and Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta psychologica*, 106(1-2):3–22.

Brass, M., Bekkering, H., Wohlschläger, A., and Prinz, W. (2000). Compatibility between observed and executed finger movements: comparing symbolic, spatial, and imitative cues. *Brain and cognition*, 44(2):124–143.

Brunetti, R., Del Gatto, C., and Delogu, F. (2014). ecorsi: implementation and testing of the corsi block-tapping task for digital tablets. *Frontiers in psychology*, 5:939.

Byrnes, J. P. (2010). Cognitive development. *The Corsini Encyclopedia of Psychology*, pages 1–3.

Cacciamani, L. and Likova, L. T. (2016). Tactile object familiarity in the blind brain reveals the supramodal perceptual-mnemonic nature of the perirhinal cortex. *Frontiers in human neuroscience*, 10:92.

Cacciamani, L. and Likova, L. T. (2017). Memory-guided drawing training increases granger causal influences from the perirhinal cortex to v1 in the blind. *Neurobiology of learning and memory*, 141:101–107.

Cao, W., Cao, X., Hou, C., Li, T., Cheng, Y., Jiang, L., Luo, C., Li, C., and Yao, D. (2016). Effects of cognitive training on resting-state functional connectivity of default mode, salience, and central executive networks. *Frontiers in aging neuroscience*, 8:70.

Capelle, C., Trullemans, C., Arno, P., and Veraart, C. (1998). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Transactions on Biomedical Engineering*, 45(10):1279–1293.

Capitani, E., Laiacona, M., Ciceri, E., et al. (1991). Sex differences in spatial memory: A reanalysis of block tapping long-term memory according to the short-term memory level. *The Italian Journal of Neurological Sciences*, 12(4):461–466.

Cappagli, G., Cocchi, E., and Gori, M. (2017a). Auditory and proprioceptive spatial impairments in blind children and adults. *Developmental science*, 20(3):e12374.

Cappagli, G., Finocchietti, S., Baud-Bovy, G., Cocchi, E., and Gori, M. (2017b). Multisensory rehabilitation training improves spatial perception in totally but not partially visually deprived children. *Frontiers in integrative neuroscience*, 11:29.

Carlesimo, G. A., Fadda, L., Lorusso, S., and Caltagirone, C. (1994). Verbal and spatial memory spans in alzheimer's and multi-infarct dementia. *Acta Neurologica Scandinavica*, 89(2):132–138.

Carreiras, M. and Codina, B. (1992). Spatial cognition of the blind and sighted: Visual and amodal hypotheses. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*.

Casey, B., Giedd, J. N., and Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological psychology*, 54(1-3):241–257.

Cattaneo, Z. and Vecchi, T. (2011). *Blind vision: the neuroscience of visual impairment*. MIT Press.

Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., and Pietrini, P. (2008). Imagery and spatial processes in blindness and visual impairment. *Neuroscience & Biobehavioral Reviews*, 32(8):1346–1360.

Caulkins, T., Corteel, E., and Warusfel, O. (2003). Wave field synthesis interaction with the listening environment, improvements in the reproduction of virtual sources situated inside the listening room. In *Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03)*.

Chakrabarti, S. and Fombonne, E. (2005). Pervasive developmental disorders in preschool children: confirmation of high prevalence. *American Journal of Psychiatry*, 162(6):1133–1141.

Collins, C. C. and Bach-y Rita, P. (1973). Transmission of pictorial information through the skin. *Advances in biological and medical physics*, 14:285–315.

Conklin, H. M., Luciana, M., Hooper, C. J., and Yarger, R. S. (2007). Working memory performance in typically developing children and adolescents: Behavioral evidence of protracted frontal lobe development. *Developmental neuropsychology*, 31(1):103–128.

Conway, A. R., Cowan, N., and Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic bulletin & review*, 8(2):331–335.

Cornoldi, C., Bertuccelli, B., Rocchi, P., and Sbrana, B. (1993). Processing capacity limitations in pictorial and spatial representations in the totally congenitally blind. *Cortex*, 29(4):675–689.

Cornoldi, C., Calore, D., and Pra-Baldi, A. (1979). Imagery ratings and recall in congenitally blind subjects. *Perceptual and Motor Skills*, 48(2):627–639.

Cornoldi, C., Rigoni, F., Venneri, A., and Vecchi, T. (2000). Passive and active processes in visuo-spatial memory: Double dissociation in developmental learning disabilities. *Brain and Cognition*.

Cornoldi, C., Tinti, C., Mammarella, I. C., Re, A. M., and Varotto, D. (2009). Memory for an imagined pathway and strategy effects in sighted and in totally congenitally blind individuals. *Acta Psychologica*, 130(1):11–16.

Cornoldi, C. and Vecchi, T. (2003). Congenital blindness and spatial mental imagery. *Y. Hatwell, A. streri & E. Gentaz (Eds.), Touching for knowing: Cognitive psychology of haptic manual perception*, pages 173–187.

Cornoldi, C. and Vecchi, T. (2004). *Visuo-spatial working memory and individual differences*. Psychology Press.

Corsi, P. M. (1973). *Human memory and the medial temporal region of the brain.* PhD thesis, ProQuest Information & Learning.

Cowan, N. and Alloway, T. (2009). Development of working memory in childhood.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., and Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of memory and language*, 31(1):1–17.

Cronly-Dillon, J., Persaud, K., and Blore, R. (2000). Blind subjects construct conscious mental images of visual scenes encoded in musical form. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1458):2231–2238.

Cuevas-Rodríguez, M., González-Toledo, D., De La Rubia-Cuestas, E., Garre, C., Molina-Tanco, L., Reyes-Lecuona, A., Poirier-Quinot, D., and Picinali, L. (2018). The 3d tune-in toolkit–3d audio spatialiser, hearing loss and hearing aid simulations. In *2018 IEEE 4th VR workshop on sonic interactions for virtual environments (SIVE)*, pages 1–3. IEEE.

Cuevas-Rodríguez, M., Picinali, L., González-Toledo, D., Garre, C., de la Rubia-Cuestas, E., Molina-Tanco, L., and Reyes-Lecuona, A. (2019). 3d tune-in toolkit: An open-source library for real-time binaural spatialisation. *PloS one*, 14(3):e0211899.

Curtis, C. E. and D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in cognitive sciences*, 7(9):415–423.

da Cunha, S. N. S., Junior, X. L. T., Guizzo, R., and de Sousa Pereira-Guizzo, C. (2016). The digital memory game: an assistive technology resource evaluated by children with cerebral palsy. *Psicologia: Reflexão e Crítica*, 29(1):5.

Davis, G., Welch, V. L., Holmes, A., and Shepherd, A. (2001). Can attention select only a fixed number of objects at a time? *Perception*, 30(10):1227–1248.

De Ajuriaguerra, J. and Stambak, M. (1969). Developmental dyspraxia and psychomotor disorders.

Denis, M., Logie, R., Cornoldi, C., de Vega, M., and Engelkamp, J. (2001). Current issues in thinking & reasoning.

Dewhurst, S. A. and Robinson, C. A. (2004). False memories in children: Evidence for a shift from phonological to semantic associations. *Psychological Science*, 15(11):782–786.

Diamond, K. E. (1990). Effectiveness of the revised denver developmental screening test in identifying children at risk for learning problems. *The Journal of Educational Research*, 83(3):152–157.

Dror, I. E. and Kosslyn, S. M. (1994). Mental imagery and aging. *Psychology and aging*, 9(1):90.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., and Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General*, 128(3):309.

Farah, M. J., Hammond, K. M., Levine, D. N., and Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive psychology*, 20(4):439–462.

Féry, Y.-A. (2003). Differentiating visual and kinesthetic imagery in mental practice. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 57(1):1.

Finke, R. A. and Freyd, J. J. (1989). Mental extrapolation and cognitive penetrability: Reply to ranney and proposals for evaluative criteria.

Flavell, J. H. (1994). Cognitive development: Past, present, and future.

Folkvord, F., Anastasiadou, D. T., and Anschütz, D. (2017). Memorizing fruit: The effect of a fruit memory-game on children's fruit intake. *Preventive medicine reports*, 5:106–111.

Foreman, N. and Gillett, R. (1997). *A handbook of spatial research paradigms and methodologies*, volume 1. Psychology Press East Sussex, UK.

Gathercole, S. and Alloway, T. P. (2008). *Working memory and learning: A practical guide for teachers*. Sage.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in cognitive sciences*, 3(11):410–419.

Gathercole, S. E., Adams, A.-M., and Hitch, G. J. (1994). Do young children rehearse? an individual-differences analysis. *Memory & Cognition*, 22(2):201–207.

Gathercole, S. E., Pickering, S. J., Ambridge, B., and Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental psychology*, 40(2):177.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., and Martin, A. J. (1999). Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 13(1):65–77.

Giannakopoulos, G., Tatlas, N.-A., Giannakopoulos, V., Floros, A., and Katsoulis, P. (2018). Accessible electronic games for blind children and young people. *British Journal of Educational Technology*, 49(4):608–619.

Gilmore, R. O. and Johnson, M. H. (1997). Body-centered representations for visually-guided action emerge during early infancy. *Cognition*, 65(1):B1–B9.

Gogtay, N., Nugent III, T. F., Herman, D. H., Ordonez, A., Greenstein, D., Hayashi, K. M., Clasen, L., Toga, A. W., Giedd, J. N., Rapoport, J. L., et al. (2006). Dynamic mapping of normal human hippocampal development. *Hippocampus*, 16(8):664–672.

Goodale, M. A. and Humphrey, G. K. (1998). The objects of action and perception. *Cognition*, 67(1-2):181–207.

Gori, M., Del Viva, M., Sandini, G., and Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, 18(9):694–698.

Gori, M., Sandini, G., Martinoli, C., and Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Current Biology*, 20(3):223–225.

Gori, M., Sandini, G., Martinoli, C., and Burr, D. C. (2013). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain*, 137(1):288–293.

Grantham, D. W. (1995). Spatial hearing and. *Hearing*, page 297.

Guttman, S. E., Gilroy, L. A., and Blake, R. (2005). Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychological science*, 16(3):228–235.

Halpern, A. R. and Zatorre, R. J. (1999). When that tune runs through your head: a pet investigation of auditory imagery for familiar melodies. *Cerebral cortex*, 9(7):697–704.

Hasher, L. and Clifton, D. (1974). A developmental study of attribute encoding in free recall. *Journal of Experimental Child Psychology*, 17(2):332–346.

Haxby, J. V., Lundgren, S. L., and Morley, G. K. (1983). Short-term retention of verbal, visual shape and visuospatial location information in normal and amnesic subjects. *Neuropsychologia*, 21(1):25–33.

Henneman, R. H. (1952). Vision and audition as sensory channels for communication. *Quarterly Journal of Speech*, 38(2):161–166.

Henry, L. (2011). *The development of working memory in children*. Sage.

Hitch, G. J., Halliday, M. S., Dodd, A., and Littler, J. E. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology*, 7(4):347–362.

Huizinga, M., Dolan, C. V., and van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, 44(11):2017–2036.

Hulme, C. and Tordoff, V. (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology*, 47(1):72–87.

Iachini, T., Ruggiero, G., and Ruotolo, F. (2014). Does blindness affect egocentric and allocentric frames of reference in small and large scale spaces? *Behavioural brain research*, 273:73–81.

Iachini, T., Sergi, I., Ruggiero, G., and Gnisci, A. (2005). Gender differences in object location memory in a real three-dimensional environment. *Brain and Cognition*, 59(1):52–59.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., and Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *science*, 286(5449):2526–2528.

Jansson, G. (1983). Tactile guidance of movement. *International journal of neuroscience*, 19(1-4):37–46.

Jeannerod, M., Arbib, M., Rizzolatti, G., and Sakata, H. (1995). Grasping objects: the cortical mechanisms. *Trends Neurosci*, 18:314–32.

Juurmaa, J. and Lehtinen-Railo, S. (1994). Visual experience and access to spatial knowledge. *Journal of Visual Impairment & Blindness*.

Kay, L. (2000). Auditory perception of objects by blind persons, using a bioacoustic high resolution air sonar. *The Journal of the Acoustical Society of America*, 107(6):3266–3275.

Kearney, G. and Doyle, T. (2015). An hrtf database for virtual loudspeaker rendering. In *Audio Engineering Society Convention 139*. Audio Engineering Society.

Kemps, E., De Rammelaere, S., and Desmet, T. (2000). The development of working memory: Exploring the complementarity of two models. *Journal of Experimental Child Psychology*, 77(2):89–109.

Kennedy, J. M. (1983). What can we learn about pictures from the blind? blind people unfamiliar with pictures can draw in a universally recognizable outline style. *American Scientist*, 71(1):19–26.

Kennedy, J. M. and Juricevic, I. (2006). Foreshortening, convergence and drawings from a blind adult. *Perception*, 35(6):847–851.

Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology*, 45(4):819.

Kessels, R. P. and Postma, A. (2018). The box task: A tool to design experiments for assessing visuospatial working memory. *Behavior research methods*, 50(5):1981–1987.

Kessels, R. P., van Den Berg, E., Ruis, C., and Brands, A. M. (2008). The backward span of the corsi block-tapping task and its association with the wais-iii digit span. *Assessment*, 15(4):426–434.

Kessels, R. P., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., and De Haan, E. H. (2000). The corsi block-tapping task: standardization and normative data. *Applied neuropsychology*, 7(4):252–258.

King, A. J., Parsons, C. H., and 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.

Kitagawa, N. and Ichihara, S. (2002). Hearing visual motion in depth. *Nature*, 416(6877):172.

Klatzky, R. L. and Lederman, S. J. (1995). Identifying objects from a haptic glance. *Perception & psychophysics*, 57(8):1111–1123.

Klingberg, T. (2006). Development of a superior frontal–intraparietal network for visuo-spatial working memory. *Neuropsychologia*, 44(11):2171–2177.

Kobayashi, M., Takeda, M., Hattori, N., Fukunaga, M., Sasabe, T., Inoue, N., Nagai, Y., Sawada, T., Sadato, N., and Watanabe, Y. (2004). Functional imaging of gustatory perception and imagery:“top-down” processing of gustatory signals. *Neuroimage*, 23(4):1271–1282.

Kosslyn, S. M. (1980). *Image and mind*. Harvard University Press.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., Hamilton, S. E., Rauch, S. L., and Buonanno, F. S. (1993). Visual mental imagery activates topographically organized visual cortex: Pet investigations. *Journal of Cognitive Neuroscience*, 5(3):263–287.

Kosslyn, S. M., Ganis, G., and Thompson, W. L. (2001). Neural foundations of imagery. *Nature reviews neuroscience*, 2(9):635–642.

Kuo, C.-Y., Huang, Y.-M., and Yeh, Y.-Y. (2018). Let’s play cards: multi-component cognitive training with social engagement enhances executive control in older adults. *Frontiers in psychology*, 9:2482.

Kwon, H., Reiss, A. L., and Menon, V. (2002). Neural basis of protracted developmental changes in visuo-spatial working memory. *Proceedings of the National Academy of Sciences*, 99(20):13336–13341.

Lay-Ekuakille, A. and Mukhopadhyay, S. C. (2010). *Wearable and autonomous biomedical devices and systems for smart environment*. Springer.

LeBoutillier, N. and Marks, D. F. (2003). Mental imagery and creativity: A meta-analytic review study. *British Journal of Psychology*, 94(1):29–44.

Lederman, S. J. and Klatzky, R. L. (1990). Haptic classification of common objects: Knowledge-driven exploration. *Cognitive psychology*, 22(4):421–459.

Lee, K., Jeong, D., Schindler, R. C., Hlavaty, L. E., Gross, S. I., and Short, E. J. (2018). Interactive block games for assessing children's cognitive skills: Design and preliminary evaluation. *Frontiers in pediatrics*, 6:111.

Lehnert, G. and Zimmer, H. D. (2006). Auditory and visual spatial working memory. *Memory & cognition*, 34(5):1080–1090.

Levtov, Y., Picinali, L., D'Cruz, M., and Simeone, L. (2016). 3d tune-in: The use of 3d sound and gamification to aid better adoption of hearing aid technologies. In *Audio Engineering Society Convention 140*. Audio Engineering Society.

Likova, L. T. (2012). Drawing enhances cross-modal memory plasticity in the human brain: a case study in a totally blind adult. *Frontiers in human neuroscience*, 6:44.

Likova, L. T. (2015). Temporal evolution of brain reorganization under cross-modal training: insights into the functional architecture of encoding and retrieval networks. In *Human Vision and Electronic Imaging XX*, volume 9394, page 939417. International Society for Optics and Photonics.

Likova, L. T. and Cacciamani, L. (2018). Transfer of learning in people who are blind: Enhancement of spatial-cognitive abilities through drawing. *Journal of Visual Impairment & Blindness*, 112(4):385–397.

Likova, L. T., Tyler, C. W., Cacciamani, L., Mineff, K., and Nicholas, S. (2016). The cortical network for braille writing in the blind. *Electronic Imaging*, 2016(16):1–6.

Logie, R. H. and Marchetti, C. (1991). Visuo-spatial working memory: Visual, spatial or central executive? In *Advances in psychology*, volume 80, pages 105–115. Elsevier.

Loomis, J. M., Lippa, Y., Klatzky, R. L., and Golledge, R. G. (2002). Spatial updating of locations specified by 3-d sound and spatial language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(2):335.

Luciana, M. (2003). The neural and functional development of human prefrontal cortex. *The cognitive neuroscience of development*, pages 157–179.

Luciana, M., Conklin, H. M., Hooper, C. J., and Yarger, R. S. (2005). The development of nonverbal working memory and executive control processes in adolescents. *Child development*, 76(3):697–712.

Luciana, M. and Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four-to eight-year-old children. *Neuropsychologia*, 36(3):273–293.

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., and Sweeney, J. A. (2004a). Maturation of cognitive processes from late childhood to adulthood. *Child development*, 75(5):1357–1372.

Luna, C., del Puerto Morales, M., Serna, C. J., and Vázquez, M. (2004b). Exchange anisotropy in co80ni20/oxide nanoparticles. *Nanotechnology*, 15(4):S293.

Maltini, E., Torreggiani, D., Venir, E., and Bertolo, G. (2003). Water activity and the preservation of plant foods. *Food Chemistry*, 82(1):79–86.

Mandler, J. M. and Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 3(4):386.

Marmor, G. S. (1978). Age at onset of blindness and the development of the semantics of color names. *Journal of experimental child psychology*, 25(2):267–278.

Martinkauppi, S., Rämä, P., Aronen, H. J., Korvenoja, A., and Carlson, S. (2000). Working memory of auditory localization. *Cerebral cortex*, 10(9):889–898.

Meijer, P. B. (1992). An experimental system for auditory image representations. *IEEE transactions on biomedical engineering*, 39(2):112–121.

Middlebrooks, J. C. (2002). Auditory space processing: here, there or everywhere? *Nature neuroscience*, 5(9):824.

Miller, N. (2015). *Dyspraxia and its Management (Psychology Revivals)*. Psychology Press.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British medical bulletin*.

Milner, D. and Goodale, M. (2006). *The visual brain in action*, volume 27. OUP Oxford.

Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 298(1089):85–95.

Miyake, A. (1999). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues, and future research directions. *Models of working memory: Mechanisms of active maintenance and executive control*, pages 442–481.

Miyake, A. and Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.

Moore, B. C. (2012). *An introduction to the psychology of hearing*. Brill.

Morrongiello, B. A., Fenwick, K. D., and Chance, G. (1998). Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events. *Infant Behavior and Development*, 21(4):543–553.

Nelson, R. E., Dickson, A. L., and Baños, J. H. (2000). An automated administration of corsi's block-tapping test. *Perceptual and motor skills*, 91(2):578–580.

Newcombe, N. S. (2002). The nativist-empiricist controversy in the context of recent research on spatial and quantitative development. *Psychological Science*, 13(5):395–401.

Oberauer, K., Schulze, R., Wilhelm, O., and Süß, H.-M. (2005). Working memory and intelligence–their correlation and their relation: comment on ackerman, beier, and boyle (2005).

Orsini, A., Chiacchio, L., Cinque, M., Cocchiaro, C., Schiappa, O., and Grossi, D. (1986). Effects of age, education and sex on two tests of immediate memory: a study of normal subjects from 20 to 99 years of age. *Perceptual and motor skills*, 63(2):727–732.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 45(3):255.

Pascual-Leone, A. and Hamilton, R. (2001). The metamodal organization of the brain. In *Progress in brain research*, volume 134, pages 427–445. Elsevier.

Pasqualotto, A., Spiller, M. J., Jansari, A. S., and Proulx, M. J. (2013). Visual experience facilitates allocentric spatial representation. *Behavioural Brain Research*, 236:175–179.

Piccardi, L., Iaria, G., Ricci, M., Bianchini, F., Zompanti, L., and Guariglia, C. (2008). Walking in the corsi test: which type of memory do you need? *Neuroscience Letters*, 432(2):127–131.

Picinali, L., Afonso, A., Denis, M., and Katz, B. F. (2014). Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge. *International Journal of Human-Computer Studies*, 72(4):393–407.

Pickering, S. J. (2001). The development of visuo-spatial working memory. *Memory*, 9(4-6):423–432.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W.-H. C., Cohen, L., Guazzelli, M., and Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences*, 101(15):5658–5663.

Ponchillia, S., Lee, H., and Moore, J. (2008). Facilitating visual efficiency, successful employment, and quality of life of working age adults with low vision (white paper). *Alexandria, VA: Association for Education and Rehabilitation of the Blind and Visually Impaired*.

Pouget, A., Deneve, S., and Duhamel, J.-R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, 3(9):741.

Požár, L. (1982). Effect of long-term sensory deprivation on recall of verbal material. *Studia Psychologica*.

Raisamo, R., Patomäki, S., Hasu, M., and Pasto, V. (2007). Design and evaluation of a tactile memory game for visually impaired children. *Interacting with Computers*, 19(2):196–205.

Renier, D., Bellato, P., Bellini, D., Pavesio, A., Pressato, D., and Borrione, A. (2005). Pharmacokinetic behaviour of acp gel, an autocrosslinked hyaluronan derivative, after intraperitoneal administration. *Biomaterials*, 26(26):5368–5374.

Rinck, M. and Denis, M. (2004). The metrics of spatial distance traversed during mental imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6):1211.

Robins, D. L., Fein, D., Barton, M. L., and Green, J. A. (2001). The modified checklist for autism in toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of autism and developmental disorders*, 31(2):131–144.

Röder, B., Rösler, F., and Neville, H. J. (2001). Auditory memory in congenitally blind adults: a behavioral-electrophysiological investigation. *Cognitive Brain Research*, 11(2):289–303.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context.* Oxford university press.

Romand, R. and Varela-Nieto, I. (2014). *Development of auditory and vestibular systems*. Academic Press.

Roncadin, C., Pascual-Leone, J., Rich, J. B., and Dennis, M. (2007). Developmental relations between working memory and inhibitory control. *Journal of the International Neuropsychological Society*, 13(1):59–67.

Röser, A., Hardiess, G., and Mallot, H. A. (2016). Modality dependence and intermodal transfer in the corsi spatial sequence task: Screen vs. floor. *Experimental brain research*, 234(7):1849–1862.

Rudkin, S. J., Pearson, D. G., and Logie, R. H. (2007). Executive processes in visual and spatial working memory tasks. *The Quarterly Journal of Experimental Psychology*, 60(1):79–100.

Ruggiero, G. and Iachini, T. (2010). The role of vision in the corsi block-tapping task: Evidence from blind and sighted people. *Neuropsychology*, 24(5):674.

Ruggiero, G., Ruotolo, F., and Iachini, T. (2009). The role of vision in egocentric and allocentric spatial frames of reference. *Cognitive processing*, 10(2):283–285.

Ruggiero, G., Ruotolo, F., and Iachini, T. (2012). Egocentric/allocentric and coordinate/categorical haptic encoding in blind people. *Cognitive processing*, 13(1):313–317.

Ruggiero, G., Sergi, I., and Iachini, T. (2008). Gender differences in remembering and inferring spatial distances. *Memory*, 16(8):821–835.

Rypma, B. and D'Esposito, M. (1999). The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. *Proceedings of the National Academy of Sciences*, 96(11):6558–6563.

Sandberg, E. H., Huttenlocher, J., and Newcombe, N. (1996). The development of hierarchical representation of two-dimensional space. *Child Development*, 67(3):721–739.

Setti, W., Cuturi, L. F., Cocchi, E., and Gori, M. (2018). A novel paradigm to study spatial memory skills in blind individuals through the auditory modality. *Scientific reports*, 8(1):13393.

Setti, W., Cuturi, L. F., Maviglia, A., Sandini, G., and Gori, M. (2019). Arena: a novel device to evaluate spatial and imagery skills through sounds. In *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE.

Shepard, R. N. and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972):701–703.

Smyth, M. M. and Scholey, K. A. (1992). Determining spatial span: The role of movement time and articulation rate. *The Quarterly Journal of Experimental Psychology Section A*, 45(3):479–501.

Sohl-Dickstein, J., Teng, S., Gaub, B. M., Rodgers, C. C., Li, C., DeWeese, M. R., and Harper, N. S. (2015). A device for human ultrasonic echolocation. *IEEE Transactions on Biomedical Engineering*, 62(6):1526–1534.

Spinnler, H. (1987). Standardizzazione e taratura italiana di test neuropsicologici. *Ital J Neurol Sci*, 6:21–120.

Squire, K. (2003). Video games in education. *Int. J. Intell. Games & Simulation*, 2(1):49–62.

St Clair-Thompson, H. and Sykes, S. (2010). Scoring methods and the predictive ability of working memory tasks. *Behavior research methods*, 42(4):969–975.

Steenari, M.-R., Vuontela, V., Paavonen, E. J., Carlson, S., Fjällberg, M., and Aronen, E. T. (2003). Working memory and sleep in 6-to 13-year-old schoolchildren. *Journal of the American Academy of Child & Adolescent Psychiatry*, 42(1):85–92.

Taevs, M., Dahmani, L., Zatorre, R. J., and Bohbot, V. D. (2010). Semantic elaboration in auditory and visual spatial memory. *Frontiers in psychology*, 1:228.

Taylor, R. L. and Warren, S. A. (1984). Educational and psychological assessment of children with learning disorders. *Pediatric Clinics of North America*, 31(2):281–296.

Thinus-Blanc, C. and Gaunet, F. (1997). Representation of space in blind persons: vision as a spatial sense? *Psychological bulletin*, 121(1):20.

Thomas, G. V. and Silk, A. M. (1990). *An introduction to the psychology of children's drawings.* New York University Press.

Thompson, P. M., Hayashi, K. M., Sowell, E. R., Gogtay, N., Giedd, J. N., Rapoport, J. L., De Zubicaray, G. I., Janke, A. L., Rose, S. E., Semple, J., et al. (2004). Mapping cortical change in alzheimer's disease, brain development, and schizophrenia. *Neuroimage*, 23:S2–S18.

Tyler, C. W. and Likova, L. T. (2012). The role of the visual arts in enhancing the learning process. *Frontiers in human neuroscience*, 6:8.

Ungar, S., Blades, M., and Spencer, C. (1995). Mental rotation of a tactile layout by young visually impaired children. *Perception*, 24(8):891–900.

Vandierendonck, A. (2000). Bias and processing capacity in the generation of random time intervals. *Cognitive Science Quarterly*.

Vandierendonck, A., Kemps, E., Fastame, M. C., and Szmalec, A. (2004). Working memory components of the corsi blocks task. *British journal of psychology*, 95(1):57–79.

Vanlierde, A. and Wanet-Defalque, M.-C. (2004). Abilities and strategies of blind and sighted subjects in visuo-spatial imagery. *Acta psychologica*, 116(2):205–222.

Vasilyeva, M. and Lourenco, S. F. (2012). Development of spatial cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(3):349–362.

Vecchi, T. (1998). Visuo-spatial imagery in congenitally totally blind people. *Memory*, 6(1):91–102.

Vecchi, T., Monticellai, M. L., and Cornoldi, C. (1995). Visuo-spatial working memory: Structures and variables affecting a capacity measure. *Neuropsychologia*, 33(11):1549–1564.

Vecchi, T., Richardson, J., and Cavallini, E. (2005). Passive storage versus active processing in working memory: Evidence from age-related variations in performance. *European Journal of Cognitive Psychology*, 17(4):521–539.

Vecchi, T., Tinti, C., and Cornoldi, C. (2004). Spatial memory and integration processes in congenital blindness. *Neuroreport*, 15(18):2787–2790.

Vingerhoets, G., De Lange, F. P., Vandemaele, P., Deblaere, K., and Achten, E. (2002). Motor imagery in mental rotation: an fmri study. *Neuroimage*, 17(3):1623–1633.

Vuontela, V., Steenari, M.-R., Carlson, S., Koivisto, J., Fjällberg, M., and Aronen, E. T. (2003). Audiospatial and visuospatial working memory in 6–13 year old school children. *Learning & Memory*, 10(1):74–81.

Wann, J., Nimmo-Smith, I., and Wing, A. M. (1988). Relation between velocity and curvature in movement: equivalence and divergence between a power law and a minimum-jerk model. *Journal of Experimental Psychology: Human perception and performance*, 14(4):622.

Warren, D. H. (1994). *Blindness and children: An individual differences approach*. Cambridge University Press.

Wimmer, M. C., Robinson, E. J., and Doherty, M. J. (2017). Are developments in mental scanning and mental rotation related? *PloS one*, 12(2).

Wright, M. (2005). Open sound control: an enabling technology for musical networking. *Organised Sound*, 10(3):193–200.

Yates, A. J. (1966). Psychological deficit. *Annual review of psychology*, 17(1):111–144.

Zimmer, H. and De Vega, M. (1996). The medium and the message in spatial communication. *Researching technical documents*, pages 153–180.