

Exploration of Architectural Spaces by Blind People using Auditory Virtual Reality for the Construction of Spatial Knowledge

Lorenzo Picinali*, Amandine Afonso**, Michel Denis, Brian FG Katz

LIMSI-CNRS, Orsay, France

Abstract

Navigation within a closed environment requires analysis of a variety of acoustic cues, a task that is well developed in many visually impaired individuals, and for which sighted individuals rely almost entirely on visual information. For blind people, the act of creating cognitive maps for spaces, such as home or office buildings, can be a long process, for which the individual may repeat various paths numerous times. While this action is typically performed by the individual on-site, it is of some interest to investigate at which point this task can be performed off-site, at the individual's discretion. In short, is it possible for an individual to learn an architectural environment without being physically present? If so, such a system could prove beneficial for navigation preparation in new and unknown environments. The main goal of the present research can therefore be summarized as investigating the possibilities of assisting blind individuals in learning a spatial environment configuration through the listening of audio events and their interactions with these events within a virtual reality experience. A comparison of two types of learning through auditory exploration has been performed: *in situ* real displacement and active navigation in a virtual architecture. The virtual navigation rendered only acoustic information. Results for two groups of five participants showed that interactive exploration of virtual acoustic room simulations can provide sufficient information for the construction of coherent spatial mental maps, although some variations were found between

*Current address: Interactive and Media Technologies research group - IMT, De Montfort University, The Gateway, LE1 9BH Leicester, UK - lpicinali@dmu.ac.uk

**Current address: Arts et Métiers ParisTech, LCPI, 75013, Paris, France

the two environments tested in the experiments. Furthermore, the mental representation of the virtually navigated environments preserved topological and metric properties, as was found through actual navigation.

Keywords: Blind people, visually impaired, spatial hearing, binaural, virtual reality, spatial cognition, room acoustic simulation

1. Introduction

Spatial audio technology has long been used within Virtual Reality (VR) applications both for recreational purposes (e.g. video games) and for studies relating to human perception, primarily in the area of auditory source localization. The ability to render individual sounds at desired positions or to create complex spatial audio scenes, without the need to manipulate any physical equipment, has offered researchers many advantages. Recently, the use of spatial audio has expanded beyond the study of such low level processes as localization, and is being used as a tool to investigate higher-level cognitive functions. The research presented in this paper results from collaboration between researchers in psychology and acoustics on the issue of spatial cognition in interior spaces.

Navigation within a closed environment requires analysis of a variety of acoustic cues, a task that is well developed in many visually impaired individuals, and for which sighted individuals rely almost entirely on visual information. For blind people, the creation of cognitive maps for spaces, such as home or office buildings, can be a long process. The individual may repeat various paths numerous times, without anybody present in the environment, in order to be able to hear all the relevant acoustic information. With this acoustic input, the individual is capable of constructing an accurate and useful mental representation of the space. While this action is typically performed by the individual on-site, it is of some interest to investigate to what degree this task can be performed off-site, at the discretion of the individual, using a virtual simulator.

As an example case, consider the situation of a blind person beginning a job in a new environment. It is common for him/her to begin by visiting the building late in the evening (when few people are present), and actively exploring the spaces in order to acquire some knowledge of the spatial configuration and of the basic features of the architectural layout and acoustic landscape (e.g. reverberation changes, background noises, sounds of foot-

steps on different floor finishes). The goal of this research is to carry out a feasibility study for investigating whether a visually impaired person would be able to collect this acoustic information – which is vital to their understanding of an environment and thus to their quality of life – by means of an interactive audio VR system which has been designed to deliver a realistic 3-D soundscape of the environment.

In short, is it possible for an individual to learn an architectural environment without being physically present? If so, such a system could prove beneficial for navigation preparation in new and unknown environments.

The motivation of the present research is therefore twofold. First, its aim is to contribute to documenting the processes by which blind people construct mental representations of their surrounding space. Secondly, it is intended to provide grounds for the design of systems delivering audio information to assist blind people in their spatial orientation and navigation of interior spaces. This is to be achieved through a feasibility study whose results will be employed for further future development and evaluations.

A comparison of two types of learning has been performed: *in situ* real displacement and active navigation in a virtual architecture. For both conditions, only acoustic information was available (participants were not allowed to use a white cane or be accompanied by a guide dog).

2. Background

This study does not present the first use of spatial audio technology and VR applications for investigating auditory performances of blind individuals. Nevertheless, the use of these technologies for exploring and evaluating spatial hearing processes and high level cognitive mechanisms can be considered as being particularly novel. This section presents previous relevant studies concerning blind people’s spatial auditory perception and use of interactive VR environments in cognitive studies.

A detailed literature review, as well as background studies related to the current work, can also be found in Katz & Picinali (2011a).

2.1. *Spatial auditory perception of visually impaired people*

It is a common belief that blind individuals have a heightened auditory sense when compared to the general sighted population. Numerous studies have been carried out in an attempt to examine this claim, focusing on different aspects such as spatial precision, response time, brain activity, and neural

plasticity. Previous studies have used either real sound sources (speakers) or Virtual Auditory Displays (VAD) with binaural renderings.

As a baseline reference, Starlinger & Niemeyer (1981) conducted a series of audiological threshold tests for 18 blind and 18 sighted individuals. Comparisons for intensity, inter-aural time difference, and auditory reflex thresholds found no significant differences. Frequency discrimination thresholds were significantly, though only slightly, improved in blind individuals.

A collection of studies focusing on spatial audition has shown a clear tendency of improved capacities in blind individuals. Focusing on central auditory system tasks, Muchnik et al. (1991) compared localization, temporal resolution, and speech extraction in noise. In comparing groups of congenital or early blind, late blind, and sighted controls (each group comprising about 10 individuals), blind participants outperformed the sighted control group. In contrast, Zwiers et al. (2001) found that elevation localization, tested in the frontal square area, deteriorated more for blind participants in low signal-to-noise conditions as compared to sighted controls, although this was tested for only a small participant population. Doucet et al. (2005) found that blind individuals had better azimuthal localization performance in the presence of spectral degradations, implying that blind individuals are better at processing spectral cues. While not specifically studying spatial audition, Röder & Rösler (2003) found that blind participants exhibited better memory for sounds, also indicating improved processing of spectral information.

Ashmead et al. (1998), in addition to providing a thorough bibliographic review on spatial audition studies, presented a study comparing Minimum Audible Angle (MAA) for azimuthal changes, showing that blind children outperformed or performed comparable to sighted adults, both being better than sighted children. Considering frontal horizontal sources, Dufour & Gérard (2000) showed that improved auditory localization performance extends to near-sighted individuals as well.

A number of studies, such as Weeks et al. (2000), have focused on evaluating neural plasticity, or changes in brain functioning, for auditory tasks executed by blind and sighted participants. Results by both Elbert et al. (2002) and Poirier et al. (2006) have shown that blind participants exhibit increased activity in areas of the brain which are typically considered visually oriented.

2.2. Human-computer interfaces for blind people

Various human-computer interfaces particularly relying on auditory and haptic stimulation have been developed and/or specifically customized for use by blind and visually impaired individuals.

In Evett et al. (2009) an interface was created to allow exploration of virtual environments via only auditory and haptic feedback using a readily available inexpensive system such as the Nintendo Wii remote. Systems such as this are meant to facilitate the development of accessible games, and thus enable universal design and accessible technologies to become more accepted.

Gomez et al. (2012) created an interface for enhancing spatial awareness and intelligibility for blind users through audio and touch. The core of this interface was the codification of color and depth information into musical instrument sounds, which provided spatial awareness and information about boundaries and obstacles. This allowed blind individuals to selectively explore environments, discover points of interest, and develop personalized navigation strategies.

Furthermore, Loomis et al. (2001) compiled a complete review of GPS-based navigation systems for visually impaired people, stressing the value of such systems for their navigation in outdoor environments. As regards indoor spaces, Loomis et al. (2005) pointed to the fact that GPS does not work fully in such contexts and that other methods are to be considered (see also Sánchez (2009)). For instance, Gill (1996) developed a solution based on infrared technology standards working as sensors to determine a traveler's indoor location. In a similar perspective, the Drishti system (Ran et al. (2004)) was designed by combining GPS for outdoor navigation and ultrasonic sensors for indoor navigation.

While outdoor navigation systems for blind people have received extensive attention in the past years (cf. Gallay et al. (2013)), indoor navigation aides are less developed, requiring further study. This is accomplished in particular through the identification of indoor wayfinding principles for blind people and the development of specific indoor routing algorithms (see Swobodzinski & Raubal (2009)). Two recent examples can be mentioned. Serrão et al. (2012) have designed a system which integrates a Geographic Information System (GIS) of a building with visual landmarks for localizing the navigating person and for tracing and validating a route for the user's navigation. Riehle et al. (2012) have reported the construction and evaluation of an indoor navigation technology intended to support seamless mobility for visually impaired people. The system infers the user's current location by

using only magnetic sensing. These are quite promising technologies. The current feasibility study however proposes a training system rather than an *in situ* navigational aid. Our approach has been developed with the intention of promoting two important features of blind people’s spatial navigation. One is the value of assisting spatial learning through the listening of acoustic events, more specifically events experienced through a virtual reality device. The other concept of the present approach is to design a system intended to assist in the preparation of a navigational episode by aiding the construction of mental spatial maps.

2.3. Interactive virtual reality environments for blind people

Numerous studies have attested to the capacity of blind people to navigate in complex environments without relying on visual inputs (Byrne & Salter (1983); Loomis et al. (1993); Millar (1994); Tinti et al. (2006); Afonso (2006); Afonso et al. (2010)). In the absence of sight, kinesthetic experience is a valid alternative source of information for constructing mental representations of an environment. A typical paradigm comprises having participants learn a new environment by locomotion (with the guidance of the experimenter), subsequently testing their performance for various mental operations on their internal representations of the environment. For instance, participants were invited to estimate distances and directions from one location to another (Byrne & Salter (1983); Veraart & Wanet-Defalque (1987); Afonso et al. (2004, 2010)).

Interesting performance patterns have been reported in studies where blind people learned the configuration of a spatial environment from verbal descriptions. Noordzij et al. (2006) compared blind and sighted people in a study in which one test involved the mental comparison of lengths between a pair of distances or line segments. Results showed that blind people were able to form spatial mental models on the basis of the descriptions, and that they exhibited the same patterns of chronometric responses as sighted people (taking more time to compare small distance differences than large ones). This effect is interpreted as reflecting that the relationships among mentally represented distances are a valid reflection of the relationships existing among actual metric distances (cf. Denis (2008); Denis & Zimmer (1992); Afonso (2006)).

Due to the absence of the visual sense, auditory sensations obviously contribute to assisting blind people in their displacements. Providing additional spatial information through audio feedback can offer valuable assis-

tance which may be implemented, for instance, in the form of an ultrasonic echolocation prosthesis (Veraart & Wanet-Defalque (1987); Wanet-Defalque & Veraart (1985)). Other interesting attempts to use audio in assistive devices are those relying on the use of virtual acoustic displays (Loomis et al. (1998); Katz et al (2012a,b)). Furthermore, several validation studies have been conducted and have shown the superiority of virtual sound for guidance when cognitive load (generated by concurrent sensory events) was present (cf. Klatzky et al. (2006); Loomis et al. (2002)).

2.4. *Localization versus spatial perception*

Sound source localization in an anechoic environment is an unusual and quite unnatural situation. It is more typical to hear sound sources with some level of reflection, even in outdoor environments, or with a high density of reflections in reverberant spaces. Depending on their intensity, these multiple acoustic reflections from the same source can cause certain impairments, such as source localization confusion and degradation of intelligibility. Quite often though, these additional acoustic signals can provide information regarding the surrounding environment, such as dimensions and material properties, as well as cues improving sound source localization.

In order to be able to localize a sound source in a reverberant environment, the human hearing system gives the most weight to the first signal that reaches the ear, i.e. the signal that comes directly from the sound source. It does not consider the localization of other signals resulting from reflections off of walls, ceiling, floor, etc. that arrive 20 – 40 ms after the first signal (these values can change depending on the typology of the signal, see Moore (2003), pp. 253-256). This effect is known as the *Precedence Effect* (Wallach et al. (1949)), and it allows for the localization of a sound source even in situations when the reflections of the sound are actually louder than the direct signal. The later arriving signals, while not being useful for localization, are used to interpret the environment.

The ability to directionally analyze the early reflection components of a sound are not thought to be common in sighted individuals for the simple reason that the information gathered from this analysis is often not needed. In fact, as already outlined in the previous sections, information about the spatial configuration of a given environment is mainly gathered through sight, and not through hearing. For this reason, a sighted individual will generally find information about the direction of the reflected signal components redundant, while a blind individual can use this information in order to gather

knowledge about the spatial configuration of an environment. Elements in support of this will be also given in Section 4.5, observing for example how blind individuals make use of self-generated noise, such as finger snaps, in order to determine the position of an object (wall, door, table, etc.) by listening to the reflections of the acoustic signals.

2.5. The proposed approach

Beyond the intrinsic value of the results of the research programs reported above, more information still needs to be collected on the conditions in which blind people use the acoustic information available in an environment to build a consistent, valid representation. It is generally recognized that the quality of such mental representations is predictive of the quality of the locomotor performance that will take place in the actual environment at some point in time. Would it be the case that a learning procedure based on systematic exploitation of acoustic cues prepares a blind person to move safely in a complex or cluttered environment? Note that the blind person who must learn a new environment in which they will have to navigate typically uses special procedures. Considering the example outlined in Section 1 (a new workplace), the objective of night-time navigation is to acquire some knowledge of the spatial configuration and of the basic features of the acoustical environment without perturbations from any other individual (including reverberation effects, sound of one’s own steps on various types of floor finish, etc.). Later on, the person will get acquainted with the daily sounds attached to every part of the environment. The research question posed in this feasibility study is whether or not a blind person could be able to collect such acoustic information – which is crucial to his/her later adaptation – via an audio VR system designed to deliver a realistic 3-D experience of the to-be-learned environment. The valuable aspect of a VR system would be to offer the possibility of conducting the acquisition phase off-site, that is, in a place other than the actual environment (for instance, the user’s home).

It is clear that most standard interactive VR systems (e.g. gaming applications) are visually-oriented. While some engines take into account source localization of the direct sound, reverberation is most often simplified and the spatial aspects neglected due to computational costs. Basic reverberation algorithms are not designed to provide such geometric information. In contrast, room acoustic auralization systems should be able to provide such level of spatial detail (see Vorländer, (2008)), though at an added computational cost. Some highly sophisticated computer games exist which render

rich and accurate auditory environments, such as *Shades of Doom* or *Lone Wolf* from GMA Games¹, but they represent an exception when looking at other gaming applications. The study presented in the following sections proposes to compare the acoustic cues provided by a real architecture with those furnished using a numerical room simulation, as interpreted by visually impaired individuals. This is seen as the first step in responding to the need of developing interactive VR systems specifically created and calibrated for blind individuals, a need that represents the principal aim of the research project discussed in the following sections.

In contrast to previous studies, this work focuses primarily on the understanding of an architectural space, and not of the sound sources in the space. Based on common situations, this study uses two corridor spaces in a laboratory building. The experiment consisted in comparing two modes of navigation along the two different corridors: a real navigation within the actual environment and a virtual navigation within an acoustical simulation of it, using a pair of headphones and a joystick. The assessment of spatial knowledge acquired in the two conditions was then examined and evaluated.

The two spaces were not exceptionally convoluted, containing a various assortment of doors, side branches, ceiling material variations, stairwells, and static noise sources. In order to enrich the soundscape within the environment itself, with the attempt of offering some additional cues to the participants for detecting, for example, the presence of obstacles, open or closed doors, etc., sound sources were added to the environments: a toilet flush noise, men's voices in a meeting, the clicking noise of a computer keyboard, and some electronic music sounds. In the real exploration condition these artificially simulated sound sources were achieved using simple audio loops played back over positioned loudspeakers. The floor plans of the spaces used in this study are shown in Fig. 1. Environment 1 was a straight corridor configuration and contained six noise sources as well as a section which ran along an open-plan office area. Environment 2 was also a straight corridor containing four noise sources and a branch junction halfway near the midpoint. Both corridors terminated with an open stairwell to one side. The difference in sound source distribution, with Environment 1 having sources throughout and Environment 2 having sources just at one end, offers two types of architectural acoustic information.

¹<http://www.gmagames.com>

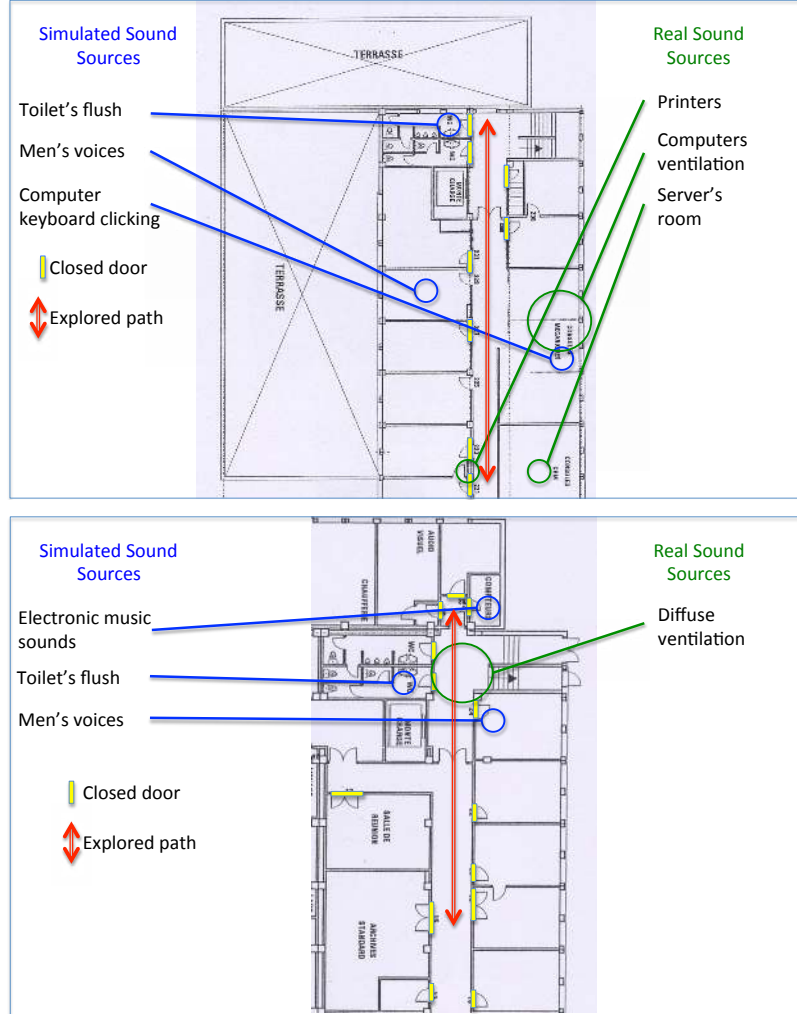


Figure 1: Plan and positions of real and artificially simulated (speaker playback of audio loops) sound sources for Environments 1 (upper) and 2 (lower). Note that for the virtual simulation of the *Computers ventilation* noise in Environment 1 two separate point sources were used to provide a somewhat diffuse source. Similarly, four point sources were employed for the virtual simulation of the *Diffuse ventilation* noise in Environment 2.

2.6. Preliminary studies

Two preliminary studies were carried out on similar topics before the one presented in this article. A brief description of these is reported below. The

basic results of these studies are provided here with appropriate references for further details.

2.6.1. Mental imagery and the acquisition of spatial knowledge without vision

The aim of this study was to investigate certain mechanisms involved in spatial cognition, with a particular interest in determining how the verbal description or the active exploration of an environment affects the elaboration of mental spatial representations. Furthermore, the role of vision was investigated by assessing whether participants without vision (blind from birth, late blind, and blindfolded sighted individuals) could benefit from these two learning modalities, with the goal of improving the understanding of the effect of visual deprivation on the capacity to mentally represent spatial configurations. Results from this earlier study suggest that active exploration of an environment enhances absolute positioning of sound sources when compared to verbal description learning. Details of this study have been presented in Afonso et al. (2005, 2010); Katz & Picinali (2011a).

2.6.2. Real vs. recorded walkthrough

The preliminary stage of the current study consisted of an evaluation of two techniques for 3D soundscape capture and restitution for a recorded walk-through exploration of the two corridor environments shown in Fig. 1. The path and, with it, the movements of the body and of the head of a blind person navigating in selected closed environments were tracked and recreated (considering also variations in the walking speed and noises generated by the individual, such as footfall sounds, finger snapping, and tongue clicking sounds), recording the experience both with a pair of binaural microphones and with a B-Format microphone.

Binaural recording is a technique which consist in recording a soundscape with a particular microphone and reproduce the audio through a pair of headphones, allowing listeners to experience a full 3D soundscape as if they were located exactly in the position of the microphone, hearing sounds as coming from the left and right, from above and below and from front and rear locations. While B-Format, and more generally Ambisonic, is a particular method which allows to record, process and synthesize 3D soundscapes which are then reproduced either through a system of loudspeakers positioned around the listener, or (as for the current study) through a pair of headphones. In the current scenario, the main difference between the two techniques is that the B-Format one allowed for head-tracking functionality,

i.e. during the playback the head movements of the participants were tracked, and the sound field was modified consequently. Differently, for the binaural recording condition when the listener moved the head during the playback, the 3D soundscape moved too, remaining in the same position relative to the head, generating an unrealistic condition.

Employing a B-Format-to-binaural conversion, with head tracking functionality, and a simple player for reproducing the binaural recordings, the two 3D audio recording techniques were compared. Using the same evaluation methodology as the current study (see Section 4.1), initial analysis quickly revealed that none of the five participants who listened to the pre-recorded audio walks were able to understand the spatial configuration of the environment under either recording condition, as compared to those individuals who performed real explorations of the same space, who were able to create accurate reconstructions. They were unable to identify the space as a corridor or a room, and they were unable to correctly identify other elements such as ceiling heights, the position of walls and obstacles, etc. This validated the initial observation, performed just after the playback task, that after listening to the 3D recorded walkthrough the participants achieved a rather poor representation of the navigated environments.

These results using spatial recordings outline how the simulation of navigation through the simple reproduction of signals recorded during a real navigation could not be considered an adequate and sufficiently precise method for the creation of a mental image of a given environment. Discussions with various participants suggested that the missing element seemed to be the lack of interactivity and freely controllable displacement within the simulated environment. Details of this preliminary study can be found in Katz & Picinali (2011a,b).

3. Virtual navigation of architectural environments

Informed by the preliminary study’s results, the current feasibility study was developed with the objective of delivering information about the spatial configuration of a closed environment, through the use of interactive virtual acoustic models. Instead of comparing real navigation with listening to spatial audio recordings, two 3D interactive acoustical models were created simulating the two real environments and the following experiment was carried out, comparing the mental representations of the environment configurations after real and virtual navigation.

3.1. Acoustic model

Using the CATT-Acoustics software², 3D architectural acoustic models were created for the two environments shown in Fig. 1. The geometrical model of Environment 1 is shown in Fig. 2. This allowed us to digitally simulate the acoustical conditions of the two environments, and to play back virtual sounds as if these were reproduced inside the environment, at a given sound source and listener position.

Room Impulse Response (RIR) measurements were performed at two positions for each environment (one position in the middle and one at the far end, near the staircase). The simulation’s material definitions were adjusted so as to obtain the same room acoustical parameters, the reverberation time calculated in octave bands, between the simulated and measured RIR in order to minimize the differences between the real and the virtual environments.

A third, more geometrically simple model was created for a training phase in order for participants to become familiar with the interface and protocol.

Previous observations in the real navigation phase highlighted that blind individuals made extensive use of self-produced noises, such as finger snapping and footsteps, in order to determine the position of an object (wall, door, table, etc.). This is possible through the perception of details and variations while listening to acoustic reflections (see Section 2.4). As such, simulation of these self-produced noises was included.

Considering the numerous sources and listener position, it was necessary to compute a large number of spatial impulse responses for the virtual active navigation rendering. A 2nd-order Ambisonic rendering engine was used, as opposed to the preliminary study’s prerecorded walkthrough which used a 1st-order Ambisonic microphone. The difference between 1st-order and 2nd-order Ambisonic is to be found in the number of channels employed (4 versus 9), resulting in an improvement of the spatial accuracy, while still allowing for dynamic head rotation of off-line rendered material.

As the room acoustic simulation software is complex, and not oriented towards real-time operation, it was necessary to pre-calculate all possible listener and source pair position combinations. Consider for example Environment 1: seven fixed sound sources had to be simulated (2 sources were used for simulating the computer ventilation noise), together with three moving ones (one for the left foot, one for the right foot, and one for the finger snap-

²<http://www.catt.se>

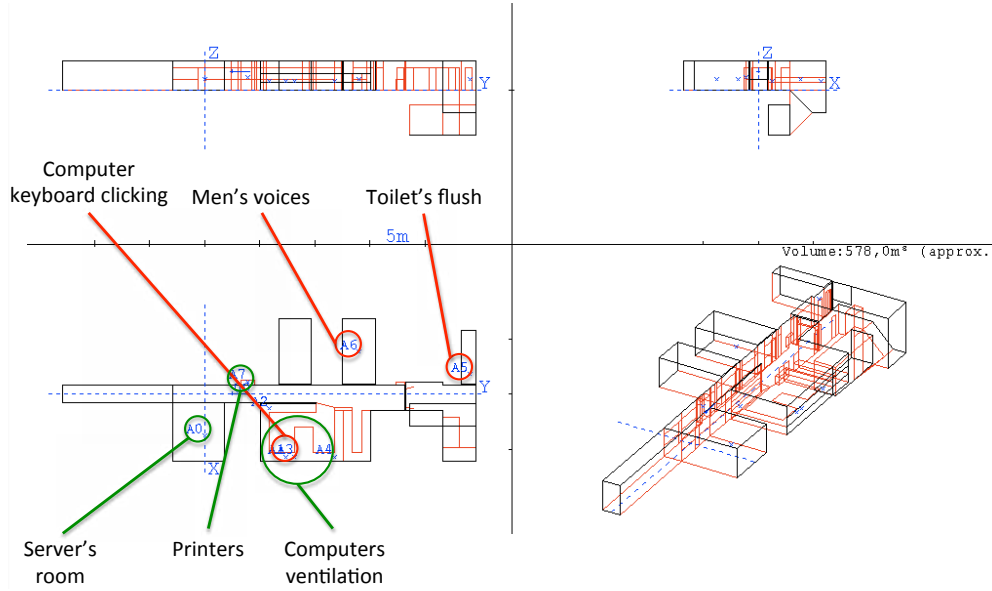


Figure 2: Geometrical acoustic model of Environment 1 including positions of real and artificially simulated sources taken from the actual environment condition.

ping), plus a moving receiver for all possible listening positions. A combined off-line/real-time approach, described in the following section, was developed for accommodating the various numbers of sources and receiver positions in the current study.

3.2. The VR navigation platform

Due to the large number of concurrent sources and the size of 2nd-order Ambisonic RIRs synthesized within the environment (resulting in an exponentially higher computation load if compared with shorter RIRs), a real-time convolution-based accurate rendering was not feasible. In order to allow for the listener to move within the virtual environment and interactively listen to the 3D soundscape, a more economical yet high performance hybrid method was developed.

As a first step, navigation was limited to one dimension only. Benefiting from the fact that both environments were corridors, the user was restricted to movements along the center-line. Receiver positions were defined at equally spaced positions (every 50 cm) along this line, at head height. The

different noise source positions, as indicated in Fig. 2, were included, providing a collection of 2nd-order Ambisonic RIR for each receiver position. In order to provide real-time navigation of such complicated simulated environments, a pre-rendering of the 2nd-order Ambisonic signals for each position of the listener was performed off-line using the audio file loops that were used in the real condition (and preliminary *in situ* recordings). This meant that for each listener's position the 3D soundscape was being continuously played back, and that at navigation time a simple 1-dimensional linear Ambisonic panning was performed between the nearest points along the center-line pathway, in order to allow the listener to hear the soundscape corresponding to his/her position in the space.

To include self-generated noises, source positions at ground level (for footfall noise) and waist height (finger snap noise) were also calculated. Spatialized finger snap and footfall noises were rendered off-line for every possible position and played back in real-time, mixed into the final Ambisonic soundscape.

In total, 44 receiver positions were calculated for Environment 1 and 31 for Environment 2. As can be seen in Fig. 2, for the Environment 1 four virtual sound sources were created for simulating the real sources in the real environment (two sources were used for simulating the computer ventilation noise), while an additional three virtual sources were created for artificially simulated sources. Similarly for Environment 2, four real sources (all used to simulate the diffuse ventilation noise) and three artificial ones were defined. In both virtual spaces a total of seven 2nd-order Ambisonic source-receiver pair RIRs were calculated for each receiver position (308 and 217 RIRs for Environments 1 and 2 respectively). In addition, for each receiver position, a corresponding RIR was calculated for simulating the finger snapping noise: the source, in this case, was different for each receiver, positioned at a height of 110 cm and a distance of 50 cm ahead of the receiver in order to more accurately represent the position of the hand's location. Finally, to account for the footstep noise, a RIR was synthesized for each receiver position at a height of 1 cm, and at a distance of 10 cm to the left of the centerline for the left step, and correspondingly to the right for the right step. The step side was alternated, starting with the right foot forward. A total of 396 2nd-order Ambisonic RIRs were synthesized for Environment 1 and 279 for Environment 2.

As outlined above, in order to provide real-time navigation of such complicated simulated environments each RIR was pre-convolved with the cor-

responding audio source signal. Recordings were made for the real sound sources present in the corridors with the microphone placed as close as possible to the noise sources in order to minimize acoustical contributions of the room in the recordings. For sources that were artificially simulated in the real environments by playing audio files over loudspeakers, the same audio files were retained. Two audio samples were selected for the finger snap and footstep noise, allowing for source variation. A total of 3564 and 2511 signals were convolved for Environments 1 and 2 respectively.

The convolved 9-channel 2nd-order Ambisonic signals corresponding to the seven static sources were summed for each receiver position. Each of the resulting mixes was then read in a customised software platform, created using the *MaxMSP* programming environment, and played back synchronously in a loop (the length of the signals was approximately 2 minutes). In order to make the processing more efficient, the multichannel player only played the signals corresponding to the two receiver positions closest to the actual position of the participant during the virtual navigation. A cosine-based cross-fade was performed between these two 2nd-order Ambisonic signals relative to the position. The playback of the convolved signals of the finger snapping noise was activated when the individual pressed one of the buttons on the joystick, cross-faded in a similar fashion. The footstep noise, with the position chosen relative to the current navigation position, was played at every displacement interval of 50 cm without any cross-fade. The choice of not having an automatic finger snapping noise, for example every 50 cm as for the footsteps, is justified considering that the extent to which these are used in real situations varies considerably between individuals, and it was desired to maintain such flexibility.

3.3. 2nd-order Ambisonic to binaural conversion

The resulting 2nd-order Ambisonic 9-channel audio stream, comprising the sum of the static sources, finger snapping, and footstep noise, was rendered binaurally over headphones employing the approach of *virtual speakers* (see McKeag & McGrath (1996); Noisternig et al. (2003)). The conversion from Ambisonic to stereo binaural signals was realized through the development and implementation of an additional *MaxMSP* software module, employing a head orientation tracking device (*XSens MTi*). The 9-channel synthesized/mixed 2nd-order Ambisonic signal was processed in real-time, performing rotations in the Ambisonic domain as a function of the participant's head orientation. The rotated signal was then decoded on a virtual

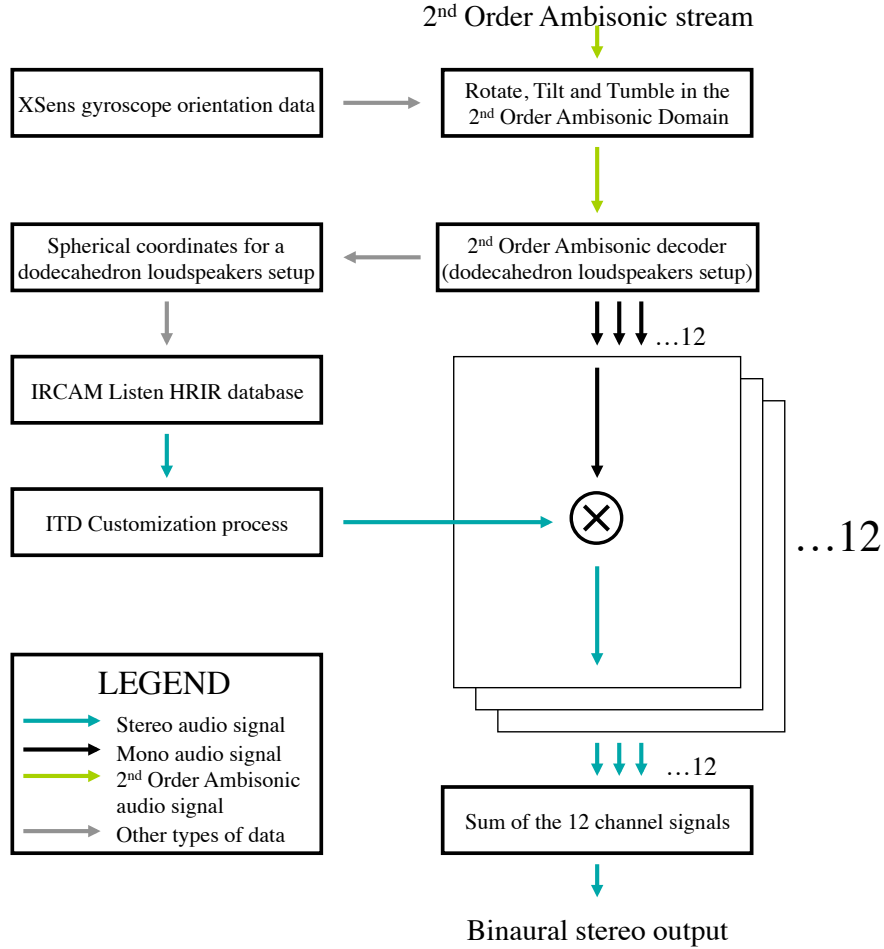


Figure 3: Schematic representation of the 2nd-order Ambisonic to binaural conversion algorithm.

loudspeaker system with the sources placed on the vertices of a dodecahedron, at 1 m distance around the center. These 12 decoded signals were then rendered as individual binaural sources via twelve instances of a binaural spatialization algorithm, converting each of the 12 monophonic signals to stereophonic binaural signals (see schematic architecture in Fig. 3). The 12 binauralized virtual loudspeaker signals were then summed and rendered to the participant, providing the complete virtual auditory environment.

The binaural spatialization algorithm used was based on the convolution

of the signal to be spatialized with a Head Related Impulse Response (HRIR) extracted from the IRCAM Listen HRTF database³. More information about this approach can be found in McKeag & McGrath (1996). The high performance spatialization engine provides binaural 3D audio rendering over headphones (Katz et al. (2010)). Full-phase HRIRs were employed, rather than minimum-phase simplifications, in order to maintain the highest level of spatial information, with a customization of the Inter-aural Time Differences (ITD), given the head circumference of the tested participant (Aussal et al. (2012)). An HRTF selection phase was also performed by each individual so that an optimal binaural conversion could be performed (see Katz & Parseihian (2012) for more details on this procedure).

In the experimental condition, participants were provided with a joystick as a navigation control device and a pair of headphones equipped with a head-tracking device. Footfall noise was automatically rendered in accordance with the participant's displacement in the virtual environment, approximating a 50 cm stride. The navigation speed was continuously variable from 0.1 to 1 m/s, proportional to the degree of forward pressure applied to the joystick. The finger snap was played each time the joystick button was pressed. Fig. 4 shows the experimental setup.

3.4. *Development summary*

In order to summarize and clarify/simplify the development exercise described in the previous sections, the following list can be considered:

- In order to simulate digitally the acoustical conditions of the two environments, and to play back virtual sounds as if these were reproduced inside the environment, two 3D architectural acoustic models were created using the CATT-Acoustics software.
- Due to the complexity of the rendered environments, in order to allow for the listener to move within the virtual environment and interactively listen to the 3D soundscape, a high performance hybrid method was developed. Navigation was limited to one dimension, and the movement was implemented through panning between pre-computed positions within the environment.

³<http://recherche.ircam.fr/equipements/salles/listen/>



Figure 4: Blind individual performing virtual navigation using the joystick and wearing headphones equipped with orientation sensor. The screen provided feedback for the operator.

- The rendered signals were then reproduced through a pair of headphones employing a customized Ambisonic to binaural conversion algorithm, which included real-time head-tracking during playback.

4. Comparison of real and virtual navigation

4.1. Protocol

The experimental evaluation consisted of comparing two modes of navigation (real and virtual) along the two different corridor environments, with the possibility offered to the participants to go back and forth along the path as many times as desired, until they were sure to have understood the spatial configuration of the environment itself. Within each environment, a number of sound sources were placed at specific locations with both sounds and their positions being the same in the real and virtual navigation conditions.

Participants were informed that the goal of the experiment was to examine to what degree, starting from the real or virtual navigation, participants could create a mental representation of the environment. They were therefore asked to try to mentally represent the configuration of the environment very accurately, trying to consider as many of the large number of elements,

and relative positions, as possible (changing floor surface materials, openings, windows, doors, objects, obstacles, etc.). They were also told that at the end of the navigation they would be asked to reconstruct the navigated environment using LEGO® blocks, outlining the various characteristics of the space, as well as the various elements/objects within it (see Section 4.4).

The reconstruction task for a given environment was performed immediately after the navigation episode within that environment.

4.2. Participants

In the real condition, two congenitally blind and three late blind individuals (three females, two males) participated in the experiment, while in the virtual condition, four early blind and one late blind individuals (three females, two males) explored the same two environments. All participants were between 24 and 57 years of age. Blindness in EB (Early Blind) participants was due to either glaucoma or congenital retinitis pigmentosa, or an excess of oxygen in the incubator. Blindness in LB (Late Blind) participants occurred between 10 and 21 years of age, and was caused by atrophy of the optical nerve, congenital cataract, glaucoma, bilateral eye tumor, or due to injury. For participants of both groups, no residual sight was present. Specific information about each participant who performed the experiment can be found in Table 1.

The participants of this study had already participated in previous experiments concerning spatial cognition, including Afonso et al. (2004, 2005); Afonso (2006); Afonso et al. (2010). Their ability to mentally represent a navigated environment was previously demonstrated, as well as the accuracy in terms of spatial configuration and the preservation of metric characteristics of objects within a mental scene.

For both real and virtual conditions, a single navigation trajectory consisted of going from the start point, to the far end, and returning. Participants were assigned to one or the other condition on a random basis, and were instructed to acquire a clear understanding of the spatial configuration of the environment. This condition had to be met before the experiment could proceed to the evaluation phase. Participants were allowed to repeat the full trajectory several times, forward then return, until they clearly understood the spatial environment. In the real condition, four participants required only a single trajectory pass, while one late blind participant needed to perform the trajectory twice. In the virtual condition, the four early blind

Condition	ID	Blindness onset	Years since sight loss	Age	Sex
Real	S1	LB	15	57	M
	S2	LB	17	30	F
	S3	EB	-	24	F
	S4	LB	10	38	M
	S5	EB	-	36	F
Virtual	S6	EB	-	38	F
	S7	EB	-	58	M
	S8	EB	-	38	M
	S9	EB	-	37	F
	S10	LB	21	31	F

Table 1: Information about the participants who performed the experiment in the real and in the virtual navigation modes, including whether participants were Early (EB) or Late (LB) Blind and if the latter, how many years since the loss of sight.

participants had to repeat the trajectory twice in order to build a clear representation of the environment, while the late blind participant had to repeat it three times. In both conditions, all participants stopped several times along the trajectory, listening to elements of the surrounding environment. In both navigation conditions, all participants used active self-generated noise (i.e. fingers snapping), reporting that it was extremely useful, together with the footstep noise, for understanding the configuration of specific parts of the environment. Early and late blind individuals used this function equally, which we discuss in Section 4.5.

The assessment of acquired spatial knowledge in the two learning conditions involved a preliminary distances comparison test (Section 4.3) and the reconstruction of the environment using LEGO® blocks (Section 4.4). Regrettably, exact exploration times for each participant were not recorded during the navigation in real and virtual environments.

4.3. Distance comparison

It is known that individuals are not only capable of recreating an object or an environment starting from perceptual stimuli and/or verbal descriptions, but they are also able of carrying out effective cognitive operations on their mental representations (Bloom et al. (1996) and Denis (1996)). As a preliminary judgment of participants mental representations, a distance

comparison task was carried out for the various sources in the real/virtual environments. The distance comparison task employed was initially proposed by Denis & Zimmer (1992), and demonstrated that sighted individuals are capable of maintaining the metric relations between various elements (sound sources, attributes of the spatial configuration, etc.) within their mental representations, and that the topological organization of the environment is preserved. Since then, the *mental comparison of distances* has been typically used in studies intended to capture the topological veridicity of represented complex environments. The major finding from such studies is that when people have to decide which of two known distances is longer, the frequency of correct responses is lower and the latency of responses is greater for smaller differences. The so-called *symbolic distance effect* is taken as reflecting the analog character of the mental representations and the capacity of preserving the metrics of the actual distances (Denis (2008); Denis & Cocude (1992); Noordzij & Postma (2005)).

These studies are particularly important, and their findings are still used in the scientific community by researchers working in different formats of mental representations. This protocol for comparing distances within mental representations was recently adapted to blind individuals, showing that both late and early blind people are able to mentally represent an environment and preserve its topological organization and the metric relation between different elements (Afonso et al. (2004, 2005, 2010)).

A test involving the mental comparison of distances was carried out for both environments, after real and virtual navigation, in order to check for the accuracy of the topological and metric properties of the mental representation of the navigated environment. In addition to the starting and ending points of the trajectory (here referred to as A and E, respectively), three sound sources were distributed along each path within the two navigated environments. For the first environment, they were *keyboard clicking* (B), *men’s voices* (C), and *toilet flush* (D). For the second environment, they were *women’s voices* (B), *electronic sound* (C), and *toilet flush* (D). For each environment, the five corresponding points (A, B, C, D, E) marked potential limits of segments along the corridors. The list of all distances between pairs of points was established and the comparison task involved a subset of these distance pairs. Following the methodology of studies on the mental comparison of distances (Denis (2008); Denis & Zimmer (1992)), the pairs selected were those having one common starting point (e.g., A-B/A-C), a selection that resulted in a total of 10 distance pairs. The pairs were categorized ac-

Distance comparison						
Environment	Real			Virtual		
Distance differences	Small	Medium	Large	Small	Medium	Large
% correct answers	92.8%	97.6%	100%	83.6%	98.8%	100%
Standard Deviation	2.95	3.29	0	11.28	2.68	0

Table 2: Percentage of correct responses for distance comparisons, combined for both environments, as a function of navigation condition.

cording to the size of the difference between the two distances mentioned (a.k.a. line segment lengths). For the pairs referring to distances that differed by one unit, distance differences were classified as “small” (A-B/A-C, A-C/A-D, A-D/A-E, B-C/B-D, B-D/B-E, C-D/C-E). For the pairs differing by two units, distance differences were classified as “medium” (A-B/A-D, A-C/A-E, B-C/B-E). For the pair differing by three units, the difference was classified as “large” (A-B/A-E). Participants were presented with each pair of distances orally and were asked to indicate which was the longer of the two.

Analysis of the results focused on the frequency of correct responses. Table 2 shows the frequency of correct responses by the participants for both real and virtual navigation conditions. The results have been averaged between the two environments, and are therefore displayed considering only real/virtual navigation and small/medium/large distance differences.

Results show that even with a high level of performance for the real navigation condition, there is a confirmation of the symbolic distance effect. The probability of making a correct decision when two distances are mentally compared increased with the size of the difference. A similar trend is seen in the virtual navigation condition. Further analysis is difficult as the results are near perfect for medium distance differences and perfect for large distance differences for both conditions.

Inferential statistics were performed to identify whether the differences between groups were statistically significant. As the data sets were small and not normally distributed, exact Mann-Whitney tests were used. For the small and medium distance difference groups, results for the real navigation did not differ significantly from the virtual (MW $U = 12$, $p = 0.917$ for the small distance difference group, MW $U = 11.5$, $p = 0.841$ for the medium distance difference group). The effect size (r) was -0.032 for the small distance

differences and -0.075 for the medium distance differences.

No analysis was performed for the large distance differences, as every participant correctly answered all the distance difference comparison tasks. These results outline that the real/virtual variable cannot be considered statistically relevant, and that for all the three types of distance differences (small, medium, and large) no significant differences between the real and virtual navigation modalities were observed.

This initial evaluation tends to confirm that both physical displacement (real navigation) and active virtual navigation with a joystick in a virtual architectural acoustic environment allowed blind individuals to create mental representations which preserved the topological and metric properties of the original environment. With these results, more detailed analysis can be performed on the map reconstructions created by each participant.

4.4. Block reconstruction

After the virtual or real navigation, participants were asked to reconstruct as accurately as possible the corridor they navigated using a set of using LEGO® construction blocks and mat. They were asked to specify everything they were able to understand of the environment and its configuration: if they detected the presence of doors, and if these were open or not; if they perceived a change in the floor surface, in the textures (paintwork, carpeting), changes in the ceiling heights, open space versus long corridors, etc. In addition, if they heard some particular noise, it was to be noted and identified. Indications of any interesting/relevant characteristics of the environments were made using special textured LEGO® blocks and specified through verbal descriptions which were annotated by the experimenter on an accompanying drawing of the reconstructed map (see examples in Fig. 5)

Before requesting participants to create the representations of the navigated environments using the blocks, they were asked if they had any previous experience with LEGO® blocks, maybe when they were children, and if they had any difficulty in manipulating them. All participants responded affirmatively to the first question, and stated that they did not have any problem using LEGO® blocks to create objects and representations. Furthermore, it is noted that all participants were familiar with other tactile representation methods such as embossed road and subway maps.

Several measures were carried out using the block reconstructions: number of sound sources mentioned, number of open doors and staircases identified, number of perceived changes of the nature of the floor finish, etc. Beyond

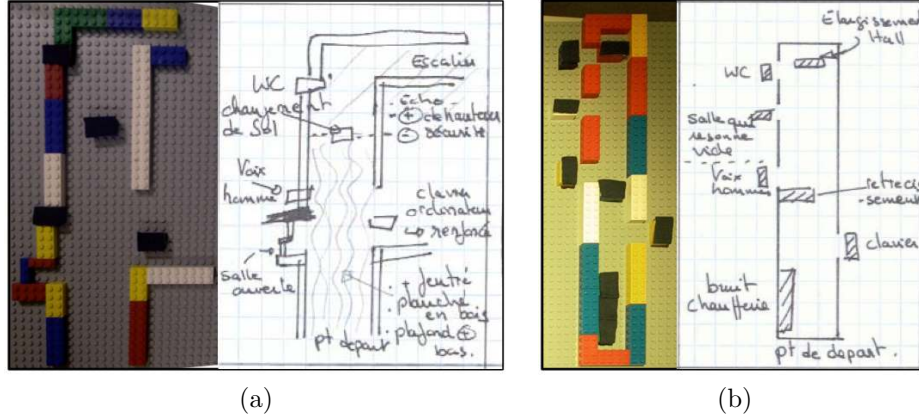


Figure 5: Examples of LEGO® reconstructions and annotated drawings for Environment 1 after real (a) and virtual (b) navigations.

some distinctive characteristics of the different reconstructions (e.g., representation of wide or narrower corridor), no particular differences were immediately evident between real and virtual navigation conditions. Both were remarkably accurate as regards the relative positions of the sound sources and acoustical architectural elements (see example in Fig. 5). Door openings into rooms containing a sound source were well identified, while more difficulty was found with those with no sound source present. Participants were also capable of distinctively perceiving the various surface material changes along the corridors.

4.4.1. Maps

Employing the participants' LEGO® maps and the experimenter's annotated drawings, the positions of the numerous reference points (architectural elements and sound sources) were manually transcribed with coordinate points in order to create numerical versions of each map. A reference template (see Fig. 6(a) for the first corridor) was created which included all sound sources and the basic architectural dimension details. A total of 48 different elements were included for the first corridor, with 47 for the second. The reference and reconstructed numerical maps for all conditions are shown in Fig. 6.

Analysis of map reconstructions can offer various metrics for performance evaluation. After navigation within a laboratory labyrinth, Passini et al. (1990) evaluated reconstructions with respect to the number of identified

elements, correct identification of said elements, and their positions. Considering the current study, a summary of the different element types identified in the various map reconstructions is provided in Tab. 3 according to Environment and navigation condition. In general, there are few identifiable differences. While the *PC Fans* were only indicated after real navigation, *Ventilation Noise* and *Printer* were only identified in the virtual condition, all being background noise-type sounds. One can note that detection of the closed elevator and open stairwell was more frequent in the real navigation condition, indicating these elements as difficult to identify in the virtual model.

A global analysis of the reconstructions relative to the reference map shows an average of $52 \pm 15\%$ and $49 \pm 17\%$ of the elements found in Environment 1 following real and virtual navigation respectively. Similar results are found for Environment 2, with $59 \pm 15\%$ and $45 \pm 15\%$ for real and virtual navigation respectively.

Inspection of the reconstruction maps highlights several points. First of all, the complexity of the maps varies between participants. For Environment 1, there is no identifiable difference between real (*S1–5*) and virtual (*S6–10*) navigation maps, while individual differences are more prominent. In contrast, the example maps from the preliminary study Katz & Picinali (2011a,b) employing binaural or Ambisonic recording playback (shown here for participants common to both studies) clearly show that participants had difficulty even correctly perceiving the basic form of the corridor. Map reconstructions for Environment 2 exhibit more simplistic models, though the general concept of the corridor and side branches were identified by some. While the Ambisonic map (*S7_{amb}*) is quite poor, it should be noted that the same individual (*S7*) generally showed the poorest performance in map reconstructions with low levels of detail.

4.4.2. Correlation Analysis

An objective evaluation on how similar the different reconstructions were from the actual maps of the navigated environments was carried out. A 2D bidimensional regression analysis (Nakaya (1997)) provided a correlation index between the reference map and each reconstructed map. This method includes some normalization in order to account for different scales between maps. Only those elements present in each individual’s reconstruction were used in the correlation computation. This resulted in a bias for maps with very few identified elements, as for example a simple rectangle could have

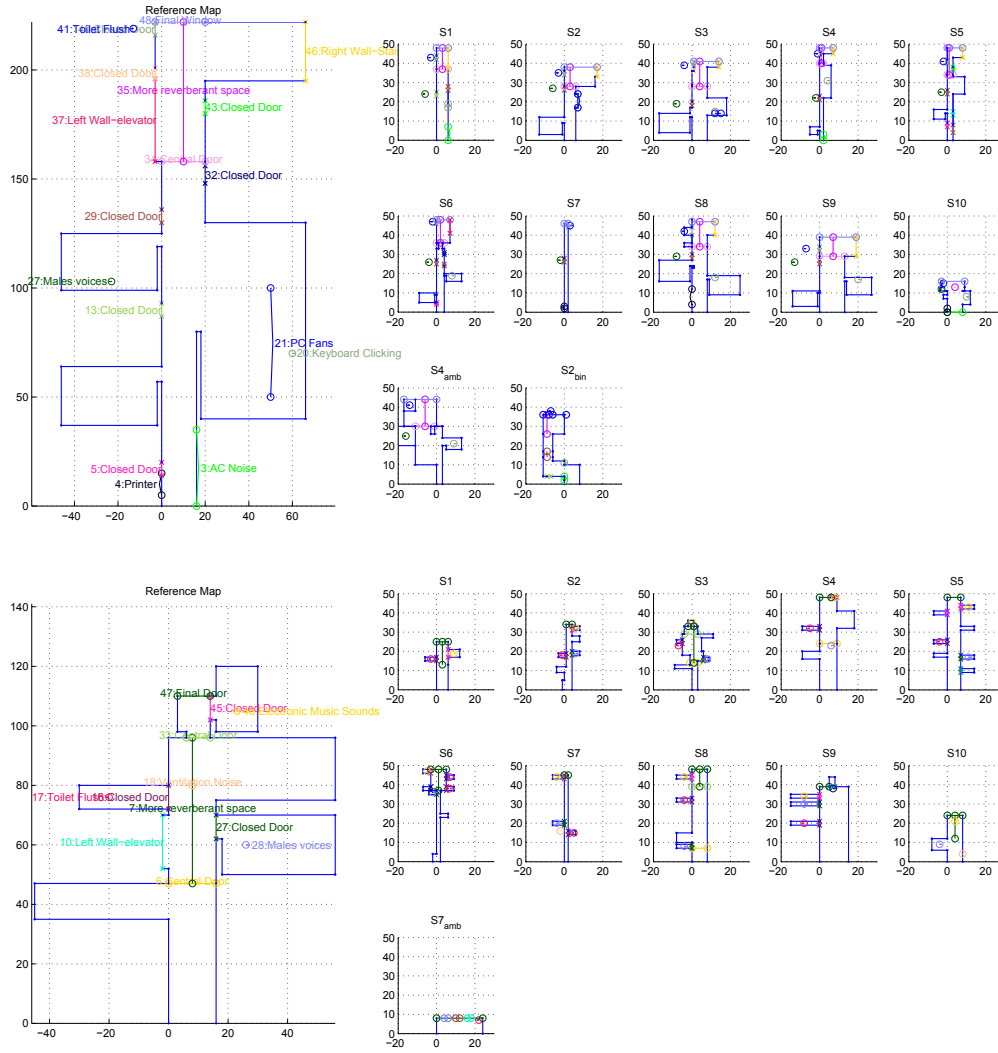


Figure 6: Reference and reconstructed maps for Environment 1 (upper) and Environment 2 (lower) after real and virtual navigation for all participants. Units in meters (reference) or LEGO® pips (reconstructions). Reconstructions for common participants following ‘Binaural’ and ‘Ambisonic’ playback passive navigation from Katz & Picinali (2011a,b) are included for comparison.

Element type	Env. 1			Env. 2		
	Real	Virtual	Actual	Real	Virtual	Actual
AirCon (<i>snd</i>)	0.8	0.4	1	0	0	1
Central Door	1.6	1.2	1	1.2	0.8	2
Closed Door	5.4	4.4	6	4.8	4.8	3
Electronic Music (<i>snd</i>)	-	-	-	1	1	1
Final Door/Window	1.6	2	1	2	2	1
Keyboard Clicking (<i>snd</i>)	0.8	0.8	1	-	-	-
Left Wall-elevator	0.4	0	1	0	0	1
Men's voices (<i>snd</i>)	0.6	1	1	0.8	1	1
More reverberant space	2	1.6	1	0.8	1.2	1
PC Fans (<i>snd</i>)	1	0	1	-	-	-
Printer (<i>snd</i>)	0	1.2	1	-	-	-
Right Wall-Stairs	1.8	0.8	1	-	-	-
Toilet Flush (<i>snd</i>)	1	1	1	1	0.8	1
Ventilation Noise (<i>snd</i>)	-	-	-	0	0.4	1

Table 3: Summary of the average number of identified elements per participant for the two Environments for real and virtual navigation conditions, including the actual number of elements in each Environment. Not all elements were present in both Environments. Sound sources are indicated (*snd*). Simple wall elements have not been included.

a high correlation index, but would not represent a high degree of understanding for the space. A summary of results for the correlation analysis of all conditions and participants is shown in Table 4. In addition to values for correlation index r^2 , values are also provided for the *Akaike Information Criterion*, AIC, which represents as a goodness-of-fit criteria according to Nakaya (1997) where the smaller the value of the AIC, the better the model.

Considering only the first environment, results for the real and virtual navigation conditions are comparable. Regarding Environment 2, the results seem to be slightly different; while the value for the real navigation condition is similar to the one of Environment 1, the correlation for the virtual condition is sensibly lower and the goodness-of-fit criteria poorer.

A more detailed presentation, regarding each map individually, is provided by comparing the correlation analysis data of all conditions and participants (see Figure 7), including the relative number of identified elements

Condition		Real	Virtual
1 st Env	Correlation index (r^2) mean	0.76	0.83
	r^2 standard deviation	0.04	0.19
	AIC mean	-138	-154
2 nd Env	Correlation index (r^2) mean	0.80	0.36
	r^2 standard deviation	0.06	0.07
	AIC mean	-173	-33

Table 4: Mean and standard deviation results for correlation index (r^2) and Akaike Information Criterion (AIC) from bidimensional regression analysis of reconstructions for both architectural environments.

in each map reconstruction. The number of participants in the current study is somewhat low, due to time required by participants and the conditions necessary for both the real and virtual navigation explorations. Despite this fact, there are a number of clearly observable tendencies. In the real exploration condition, both the correlation index and the quantity of identified elements are rather high.

As no participant performed more than one condition, there are likely to be some individual variance effects, but all participants expressed their comfort in the task and their understanding of the space at the time of the experiment. Due to the limited number of participants, no specific analysis was performed comparing the results between early and late blind individuals. As noted earlier, participant *S7* created rather simple maps. In the virtual navigation condition in Environment 1, the *S7* map reconstruction contained less than 20% of the reference elements. As such, an alternate confidence ellipse is also included with this participant removed.

There was a notable difference in the results between Environments 1 and 2, with the virtual condition in general providing similar correlation values to the real condition in Environment 1. The results for common participants from the previous preliminary study using binaural and Ambisonic playback are also included. For Environment 1 the results are comparable. Considering the visual examination of the map reconstructions in Fig. 6($S2_{bin}, S4_{amb}$), this may indicate the limitations of using this method of analysis in this situation. In Environment 2, the virtual condition was not comparable to the real condition with an overall shift in correlation results.

Inferential statistics were performed to identify whether the differences

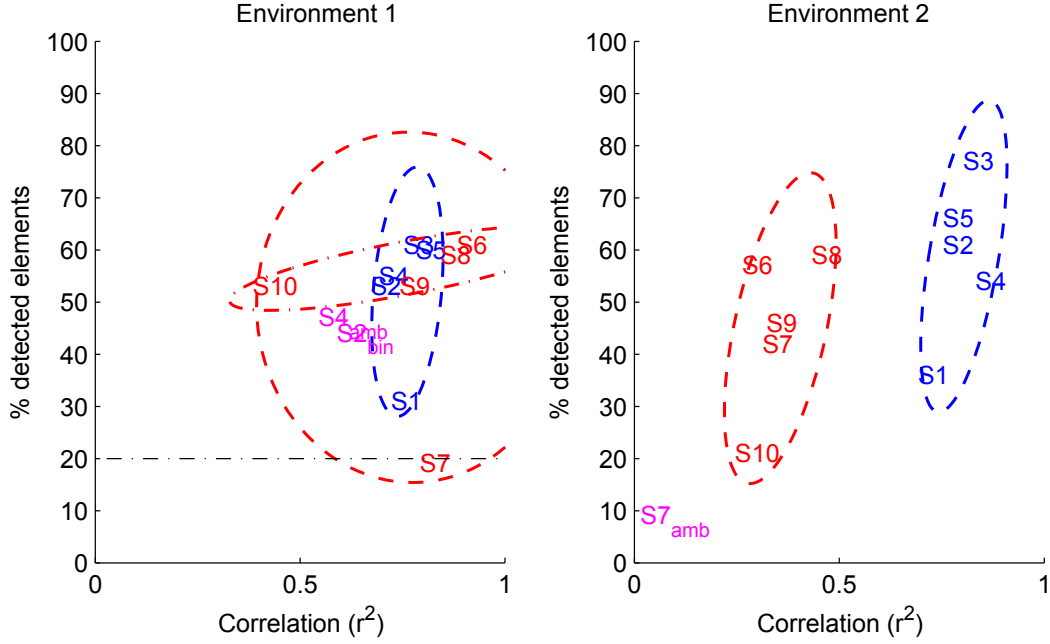


Figure 7: Correlation index versus percentage of identified elements in the map reconstruction for all participants for Environment 1 and Environment 2 including confidence ellipses real (red $--$) and virtual (blue $--$) navigation conditions. Confidence ellipse for Environment 1 excluding $S7$, who identified less than 20% elements, is shown (red $- \cdot -$). Results from the preliminary study with 3D recordings condition *ambisonic* and *binaural* for common participants are also included for comparison.

between groups were statistically significant. As the data sets were small and not normally distributed, exact Mann-Whitney tests were used. For Environment 1, correlation values for the real navigation did not differ significantly from the virtual (MW $U = 5.50$, $p = 0.175$). The effect size (r) was -0.465 . For Environment 2, correlation values for the real navigation did differ significantly from the virtual (MW $U = < 0.00$, $p = 0.008$). The effect size (r) was -0.831 . This result confirms the outcomes outlined above; that for Environment 1 no significant differences between the real and virtual conditions were observed.

For Environment 2, the statistical relevance of the real/virtual variable was higher, therefore there was a difference in terms of 2D correlation between the two conditions, with the real condition having a higher mean correlation

(approximately twice) than the virtual condition.

The results of this feasibility study could be described as having low statistical power. This suggests that differences between data sets will be under-estimated (type two error). Interestingly, the effect sizes identified suggest that there may be relevant relationships between variables, which could be explored if the research was repeated with larger data sets. It is therefore possible to say that increasing the number of participants could lead to a convergence in terms of correlation values between virtual and real navigation in the two environments.

The notable difference between the map reconstruction analysis and results for the two Environments is worth discussing. While it is not possible to conclude definitively the reasons behind the poorer performance for Environment 2 under virtual navigation, some of the main differences between the two virtual models can be noted. With regards to static position sound sources (not considering self-generated sounds) Environment 1 contained sources throughout the corridor while Environment 2 had the sources congregated at one end (see Fig. 1). Direct inspection of the map reconstructions (Fig. 6) does not necessarily show such contrast as the correlation analysis provides. This may lead to some doubt about the application of this analysis method to the map generation protocol in the current experiment. The fact that participants were relatively free in their creation possibly led to variations in level of detail between individuals, as well as individual emphasis on different elements. Subsequent studies should attempt to provide stricter guidelines in order to produce maps which are more comparable in their content and method of construction.

4.5. Participant feedback

Some points of interest were reported by the participants in the virtual navigation condition. Many mentioned explicitly that for sound sources that were located at the left or at the right of the corridor, they perceived both the direct sound coming from the source and the reflected sound coming from the opposite direction (reflection off the wall), making it possible to locate both the source on one side and the reflecting object on the other. The presence of both dynamic cues relative to displacement and controlled events such as finger snaps, as included in the virtual condition, were deemed highly valuable by the participants, and the virtual acoustic simulations provided acoustic information that allowed for highly correlated detailed map reconstructions relative to a real exploration condition.

The finger snap sound (auditory feedback) was widely used by all participants; they all pointed out that listening to the reverberation generated by the noise was helpful in understanding the spatial configuration and the structure of various parts of the navigated environments. Furthermore, the footstep noise, and the possibility of continuously varying the navigation speed, were found to be very informative. These factors can be considered as important in light of what has been described in Section 2.4.

The results of this feasibility study are promising with respect to the hypothesis that exploration within an interactive virtual acoustical environment, as developed here, is sufficiently realistic and well-defined to offer suitable information for spatial understanding of the architecture. An interactive virtual simulation can be precise enough so that information about the spatial configuration of the total environment, not just source positions, can be gathered by visually impaired individuals solely through auditory exploration. However, additional studies are necessary to determine which elements are required in the creation of the virtual environment to ensure such an understanding due to the fact that the two test environments did not offer the same results.

5. Conclusions

In the studies presented in this paper, the possibility of assisting blind individuals in learning the configuration of a closed space through 3D audio recording and simulation techniques has been investigated. Interpretations of the presented acoustic information through block map reconstructions were compared to reconstructions following *in situ* exploration. Previous work considered exploration via playback of binaural and Ambisonic walk-through recordings of the same spaces.

A preliminary study (Section 2.6.2) attested that passive listening to binaural recordings was not efficient, presumably because of the absence of displacement and the absence of head orientation information. While offering head tracking possibilities, 1st-order Ambisonic recordings were less satisfactory than binaural recordings due to poor localization accuracy, in addition to the absence of displacement control and displacement information. In the present study (Section 3), interactive VR navigation performed better, by including both head tracking and controlled displacement. These results are of special value as no significant difference was found for behavioral measures between real and virtual navigation conditions. Using joystick controlled

navigation, it was possible for participants to build mental representations of the spatial structure of indoor environments which included both sound sources and realistic room acoustic simulations. This mental representation preserved the topological and metric properties of the environment, as was found through actual navigation.

5.1. Guidelines for the design of VR systems for visually impaired people

The outcomes from this feasibility study could benefit the development of acoustic interactive VR systems to give visually impaired individuals the possibility of gathering information about the spatial configuration of a building and its environments, to become aware of the obstacles and architectural barriers, and to become familiar with a future route whilst remaining seated at home, simply interacting with the computer.

Consider for example a visually impaired person who plans to go to a concert, and purchases an online ticket. The individual has not visited this particular concert hall before and has no prior knowledge of paths to follow between the entrance of the building, the restaurant, concession stand, and his or her own seat. An interactive acoustic VR system could be included on the concert hall website to enable the individual to learn about the spatial configuration of the concert hall and its environs, including awareness of obstacles and architectural barriers, and thus to become familiar with the various potential paths of interest of the evening in advance of the actual event.

Such a system could therefore be created for concert halls, museums, municipal buildings, open spaces (e.g. university campuses), etc. The creation process should follow the procedures used for the creation of the models in the current research (e.g. see Section 3), considering also the guidelines outlined in the following paragraphs. The system could then be implemented on a client-server architecture for remote usage over the Internet, or on a standalone computer to be installed, for example, at the entrance of a museum.

There are two relevant design implications for the creation of a VR system for visually impaired individuals resulting from this study. The first one concerns interactivity; the system needs to maintain a very high level of interactivity, and this was demonstrated by the fact that the playback of 3D sounds in the preliminary study (Section 2.6.2), where there was minimal (head tracking) or no level of interaction, did not allow the participants to create a mental image of the spatial configuration of the environments. In the current VR interactive application, the possibility given to participants to use

head tracking and to dynamically choose their navigation speed were found to be extremely important by all participants in helping them understand the spatial configuration of the environments and the location of sound sources (Section 4.5).

The second design implication pertains to the activity of echolocation; the directional listening of the reflections of sound waves off of surfaces, objects, etc. This skill appears to be a highly relevant mechanism used by blind individuals for understanding the configuration of an environment, or parts of it. This skill has also been reported in sighted individuals, but is not as pertinent or as well developed as with visually impaired people (Stoffregen & Pittenger (1995)). A VR interactive system created for supporting visually impaired individuals needs to include the possibility of generating sounds and noises (e.g. finger snapping, footstep noise) in order to perceive to the acoustical reflections of the generated sounds in the virtual environment. Related to the calculation and rendering of acoustical reflections, one should avoid, or at least limit, approximations or simplifications (such as modeling only low reflection orders) affecting directional accuracy which are common in VR and gaming applications (see Section 2.5). Such simplifications, while improving computational performance, could result in major degradation issues for blind users when trying to create a mental representation of the environment and to understand its spatial configuration.

The differences between the two Environments leads to certain questions pertaining to model creation, such as the level of geometric complexity and the number and distribution of static sound sources needed for reliable comprehension of the virtual architecture. From these initial results, it is not possible to extract guidelines pertaining to these questions. However, it is hoped that future studies will benefit from the experimental design and results shown here in their detail. Further investigations should be carried out in order to examine these specific factors affecting dynamic virtual navigation.

5.2. Future developments

Possible future developments for similar applications could be linked to the design of a 3D VR acoustical model specifically customized for blind people. In the present research, generalized room acoustics prediction software used by acousticians was employed. Despite the fact that auditory perception and interpretation skills can vary between blind individuals, for example in terms of environmental comprehension and localization accuracy,

additional studies should be carried out to determine the pertinent acoustical characteristics of a simulated space. This knowledge could help reduce the computational load and complexity of the simulation, potentially allowing for real-time free navigation. Furthermore, extensive behavioral tests should be carried out, even with a small number of participants, on simpler environmental configurations, in order to study possible augmentations/reductions in the rendered architectural details, performing a “fine tuning” of the accuracy of the acoustical characteristics of the simulated spaces.

Other future developments could include the investigation of different navigation methods, comparing for example navigation using a joystick with full body and displacement tracking. Furthermore, the implementation of such applications on a client-server based platform could be examined, allowing blind individuals to access VR interactive environments directly from a typical home computer, thereby increasing the usability and societal impact of this particular research.

6. Acknowledgments

The authors would like to thank all the participants who took part in the study, and in particular Mr. Bernard Lebahr for accepting to be photographed for Fig. 4. This research was supported in part by a grant from the European Union to Michel Denis (STREP “Wayfinding”, Contract #12959). Experiments were approved by the Ethics Committee of the National Center for Scientific Research (Comité Opérationnel pour l’Ethique en Sciences de la Vie). The authors would like to thank Peter Norrie (DMU) for assisting with some statistical analysis. Final thanks are offered to Dr. Nakaya for providing the bi-dimensional regression analysis spreadsheet.

References

- Afonso, A. (2006). *Propriétés analogiques des représentations mentales de l’espace: Etude comparative de personnes voyantes et non-voyantes*, Doctoral dissertation, Université Paris Sud, Orsay, France.
- Afonso, A., Blum, A., Katz, B.F.G., Tarroux, P., Borst, G., & Denis, M. (2010). Structural properties of spatial representations in blind people : scanning images constructed from haptic exploration or from locomotion in a 3-D audio virtual environment. *Memory & Cognition*, vol. 38, pp. 591-604.

- Afonso, A., Gaunet, F., & Denis, M. (2004). *The mental comparison of distances in a verbally described spatial layout: Effects of visual deprivation, Cognition and Personality*, Special issue on Imagination, vol. 23, pp. 173-182.
- Afonso, A., Katz, B.F.G., Blum, A., Jacquemin, C., & Denis, M. (2005). A study of spatial cognition in an immersive virtual audio environment: Comparing blind and blindfolded individuals, *Proc. of the 11th Meeting of the International Conference on Auditory Display*, 6-9 July, Limerick, Ireland.
- Ashmead D.H., Wall R.S., Ebinger K.A., Eaton, S.B., Snook-Hill M-M, & Yang X. (1998). Spatial hearing in children with visual disabilities, *Perception*, 27, pp. 105-122.
- Aussal, M., Alouges, F., & Katz, B.F.G. (2012) ITD interpolation and personalization for binaural synthesis using spherical harmonics, *Proc. of the 25th Audio Engineering Society UK Conference*, 25-27 March, York, UK, pp. 04.01-10.
- Blauert, J. (1996). *Spatial Hearing: Psychophysics of Human Sound Localization*, 2nd edition, MIT Press, Cambridge, Massachussets.
- Bloom, P., Peterson, M. A., Nadel, L., & Garrett, M. F. (1996). *Language and space*. Cambridge, MA: MIT Press.
- Byrne, R. W. & Salter, E. (1983). Distances and directions in the cognitive maps of the blind, *Canadian Journal of Psychology*, 70, pp. 293-299.
- Denis, M. (1996). Imagery and the description of spatial configurations, in *Models of visuo-spatial cognition*, De Vega, Intons-Peterson, M. J., M., Johnson-Laird, P.N., Denis, M. & Marschark, M. Eds., New York: Oxford University Press, pp. 128-197.
- Denis, M. (2008). Assessing the symbolic distance effect in mental images constructed from verbal descriptions: A study of individual differences in the mental comparison of distances. *Acta Psychologica*, 127, pp. 197-210.
- Denis, M. & Cocude, M. (1992). Structural properties of visual images constructed from poorly or well-structured verbal descriptions, *Memory and Cognition*, Vol. 20, pp. 497-506.

- Denis, M. & Zimmer, H. D. (1992). Analog properties of cognitive maps constructed from verbal descriptions, *Psychological Research*, 54, pp. 286-298.
- Doucet, M.E., Guillemot, J.P., Lassonde, M., Gagné J.P., Leclerc, C., & Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals, *Experimental Brain Research*, 160, No. 2, pp. 194-202.
- Dufour, A. & Gérard, Y. (2000). Improved auditory spatial sensitivity in nearsighted subjects, *Cognitive Brain Research*, 10, pp. 159-165.
- Elbert, T., Sterr, A., Rockstroh, B., Pantev, C., Muller, M.M., & Taub, E. (2002). Expansion of the Tonotopic Area in the Auditory Cortex of the Blind, *The Journal of Neuroscience*, 22, pp. 9941-9944.
- Evetts, L., Battersby, S., Ridley, A., & Brown, D. (2009). An interface to virtual environments for people who are blind using Wii technology - mental models and navigation, *Journal of Assistive Technologies*, 3(2), pp. 26 - 34.
- Gallay, M., Denis, M., & Auvray, M. (2013). Navigation assistance for blind pedestrians: Guidelines for the design of devices and implications for spatial cognition. *Representing space in cognition: Interrelations of behaviour, language and formal models*, T. Tenbrink, J. Wiener & C. Claramunt (Eds.), Oxford, UK: Oxford University Press, pp. 244-267.
- Gerzon, M. A. (1972). Periphony With-Height Sound Reproduction, *Proc. of the 2nd Convention of the Central Europe Section of the Audio Engineering Society*, Munich, Germany.
- Gill, J. (1996). An orientation and navigation system for blind pedestrians. *MoBIC Consortium, London: Royal National Institute for the Blind*.
- Gomez, D.J., Bologna, G., & Pun, T. (2012). Spatial awareness and intelligibility for the blind: audio-touch interfaces, *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, pp. 1529-1534.
- Hammershøi, D. & Møller, H. (2002). Methods for Binaural Recording and Reproduction, *Acta Acustica united with Acustica*, 88(3), pp. 303-311.

- Katz, B.F.G., Dramas, F., Parseihian, G., Gutierrez, O., Kammoun, S., Brilhault, A., Brunet, L., Gallay, M., Oriola, B., Auvray, M., Truillet, P., Denis, M., Thorpe, S., & Jouffrais, C. (2012a) NAVIG: Guidance system for the visually impaired using virtual augmented reality. *Technology and Disability*, v. 24 (2), pp. 163-178.
- Katz, B.F.G., Kammoun, S., Parseihian, G., Gutierrez, O., Brilhault, A., Auvray, M., Truillet, P., Denis, M., Thorpe, S., & Jouffrais, C. (2012b) NAVIG: Augmented reality guidance system for the visually impaired: Combining object localization, GNSS, and spatial audio. *Virtual Reality*, 2012 vol. 16(3), pp. 253-269.
- Katz, B.F.G. & Parseihian, G. (2012) Perceptually based head-related transfer function database optimization. *Journal of the Acoustical Society of America*, vol. 131(2), pp. EL99-EL105.
- Katz, B.F.G. & Picinali, L. (2011a). Spatial Audio Applied to Research with the Blind, in *Advances in Sound Localization*, Sturmillo, P. Ed., INTECH, April 2011, ISBN: 978-953-307-224-1.
- Katz, B.F.G. & Picinali, L. (2011b). Exploration of Virtual Acoustic Room Simulations by the Visually Impaired. *Proc. of the International Seminar on Virtual Acoustics*, 24-25 November, Valencia, Spain.
- Katz, B.F.G., Rio, E., & Picinali, L. (2010). LIMSI Spatialization Engine. Inter Deposit Digital Number IDDN.FR.001.340014.000.S.P.2010.000.31235.
- Klatzky, R. L., Marston, J. R., Giudice, N. A., Golledge, R. G., & Loomis, J. M. (2006). Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *Journal of Experimental Psychology: Applied* 12, pp. 223-232.
- Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (1998). Navigation system for the blind: Auditory display modes and guidance, *Presence: Teleoperators and Virtual Environments*, 7, 193-203.
- Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (2001). GPS-based navigation systems for the visually impaired, in *Fundamentals of wearable computers and augmented reality*, Barfield, W. & Caudell, T. Ed., pp. 429-446, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

- Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., & Fry, P. A. (1993). Nonvisual navigation by blind and sighted: Assessment of path integration ability, *Journal of Experimental Psychology: General*, 122, pp. 73-91.
- Loomis, J. M., Lipka, Y., Klatzky, R. L., & 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, pp. 335-345.
- Loomis, J. M., Marston, J. R., Golledge, R. G., & Klatzky, R. L. (2005). Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance. *Journal of Visual Impairment and Blindness*, 99, pp.219-232.
- McKeag, A. & McGrath, D. S. (1996). Sound Field Format to Binaural Decoder with Head Tracking, *Proc. of the 101th Audio Engineering Convention*, Los Angeles, CA.
- Millar, S. (1994). *Understanding and representing space: Theory and evidence from studies with blind and sighted children*, Oxford, UK: Clarendon Press.
- Moore, Brian C. J. (2003). *An Introduction to the Psychology of Hearing*, Fifth Edition, London, UK: Academic Press.
- Muchnik C., Efrati M., Nemeth E., Malin M., & Hildesheimer M. (1991). Central auditory skills in blind and sighted subjects, *Scandinavian Audiology*, 20, pp. 19-23.
- Nakaya,T. (1997): Statistical inferences in bidimensional regression models. *Geographical Analysis*, Vol. 29, pp. 169-186.
- Noisternig, M., Musil, T., Sontacchi, A., & Holdrich, R. (1997). 3D binaural sound reproduction using a virtual ambisonic approach. *Proc. of the 2003 IEEE Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, 27-29 July, Lugano, Switzerland.
- Noordzij, M. L., & Postma, A. (2005). Categorical and metric distance information in mental representations derived from route and survey descriptions, *Cognition*, Vol. 100, pp. 321-342.

- Noordzij, M. L., Zuidhoek, S., & Postma, A. (2006). The influence of visual experience on the ability to form spatial mental models based on route and survey descriptions. *Cognition*, 100, pp. 321-342.
- Passini, R., Proulx, G., & Rainville, C. (1990). The spatio-cognitive abilities of the visually impaired population. *Environment and Behavior* 22, pp. 91-118.
- Poirier, C., Collignon, O., Scheiber, C., & De Volder, A. (2006). Auditory motion processing in early blind subjects, *Cognitive Processing*, 5(4), pp. 254-256.
- Ran, L., Helal, S., & Moore, S. (2004). Drishti: An integrated indoor/outdoor blind navigation system and service. *Proc. of the 2nd International Conference on Pervasive Computing and Communications*, Orlando, Florida, March 2004, pp. 23-30.
- Riehle, T. H., Anderson, S. M., Lichter, P. A., Giudice, N. A., Knuesel, R. J., Kollmann, D. T. & Hedin, D. S. (2012). Indoor magnetic navigation for the blind. *Proceedings of the 34th Annual IEEE Engineering in Medicine and Biology Conference*, Dan Diego, California, pp. 1972-1975.
- Röder B. & Rösler F. (2003). Memory for environmental sounds in sighted, congenitally blind and late blind adults: Evidence for cross-modal compensation, *Int J Psychophysiol*, 50, pp. 27-39.
- Sánchez, J. (2009). Mobile audio navigation interfaces for the blind. In *Universal Access in human-computer interaction: Intelligent and ubiquitous interaction environments (Part II)*, Berlin: Springer, pp. 402-411.
- Serrão, M., Rodrigues, J. M. F., Rodrigues, J. I. & du Buf, J. M. H. (2012). Indoor localization and navigation for blind persons using visual landmarks and GIS. *Procedia Computer Science*, 14, pp. 65-73.
- Starlinger I. & Niemeyer, W. (1981). Do the Blind Hear Better? Investigations on Auditory Processing in Congenital or Early Acquired Blindness, *Audiology*, 20, pp. 503-509.
- Stoffregen T. A. & Pittenger, J. B. (1995). Human Echolocation as a Basic Form of Perception and Action, *Ecological Psychology*, Vol. 7(3), pp. 181-216.

- Swobodzinski, M. & Raubal, M. (2009). An indoor routing algorithm for the blind: Development and comparison to a routing algorithm for the sighted. *International Journal of Geographical Information Science*, 23, pp. 1315-1343.
- Tinti, C., Adenzato, M., Tamietto, M. & Cornoldi, C. (2006). Visual experience is not necessary for efficient survey spatial cognition: Evidence from blindness, *Quarterly Journal of Experimental Psychology*, 59, pp. 1306-1328.
- Veraart, C. & Wanet-Defalque, M. C. (1987). Representation of locomotor space by the blind, *Perception and Psychophysics*, 42, pp. 132-139.
- Vorländer, M. (2008). *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Aachen, Germany: Springer-Verlag. ISBN: 978-3-540-48829-3.
- Wallach, H., Newman, E. B. & Rosenzweig, M. R. (1949). The precedence effect in sound localization, *Journal of Experimental Psychology*, Vol. 27, pp. 339-368.
- Wanet-Defalque, M. C. & Veraart, C. (1985). Processing of auditory information by blind in spatial localization task. *Perception and Psychophysics* 38, pp. 91-96.
- Weeks, R., Horwitz, B., Aziz-Sultan, A., Tian, B., Wessinger, C. M., Cohen, L.G., Hallett, M., & Rauschecker, J.P. (2000). A Positron Emission Tomographic Study of Auditory Localization in the Congenitally Blind, *J. Neuroscience*, 20, pp. 2664-2672.
- Zwiers M.P., Van Opstal A.J., & Cruysberg J.R. (2001). A spatial hearing deficit in early-blind humans, *The Journal of Neuroscience*, 21(9):RC142, pp. 1-5.