13

Spatial Audio Applied to Research with the Blind

Brian FG Katz[1] and Lorenzo Picinali[2]
[1]*Déptartement Communication Homme-Machine, LIMSI-CNRS, Orsay*
[2]*Department of Media Technology, De Montfort University, Leicester*
[1]*France*
[2]*UK*

1. Introduction

Spatial audio technology has long been used for studies relating to human perception, primarily in the area of auditory source localisation. The ability to render individual sounds at desired positions or 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 localisation, and has been used as a tool to investigate higher-level cognitive functions. This work presents several recent studies where spatial audio technology has been used in order to expand our understanding of spatial cognition, with a specific focus on the abilities of the visually impaired, in both free-field and interior space exploration. These types of works provide for both an improved understanding within cognitive science and for the research and development into improved high resolution renderings with appropriate auditory cues.

2. Spatial auditory perception of the visually impaired

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, from spatial precision to 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, interaural-time difference, and auditory reflex thresholds found no significant differences. Frequency discrimination thresholds were slightly improved, though still significantly, in blind individuals. Focusing on pitch discrimination, Gougoux et al. (2004) found that early blind individuals were better than sighted control subjects in detecting pitch changes for a pair of sinusoids.

A collection of studies focusing on spatial audition has shown a clear tendency in support 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 of about 10 individuals, blind subjects outperformed the sighted control group with the blind group localizing source positions within an error span of $\pm 5^\circ$ at a rate of $\approx 76\%$, while the sighted group had a correct response rate of ≈ 52% for the different sources tested in the horizontal plane. Temporal resolution was improved for both blind groups, and speech discrimination levels for noise masked speech were 6% higher when compared to the sighted control group. In contrast, Zwiers et al. (2001a) found that elevation localization, tested in the frontal square area of ±35 ° in azimuth and elevation, deteriorated more for blind subjects in low signal-to-noise conditions as compared to sighted control, although this was tested only for a small subject 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 subjects exhibited better memory for sounds, also indicating improved processing of spectral information.

Lessard et al. (1998) found comparable frontal lateralization between sighted and blind subjects while in the case of monaural localization (spectral discrimination) an improvement was evident for blind subjects. Examination of localization precision as a function of source location by Röder et al. (1999) found comparable results in the frontal direction and improved localization for more lateral sounds in blind subjects.

Ashmead et al. (1998), in addition to providing a thorough bibliographic study on spatial audition studies, presented a study comparing Minimum Audible Angle (MAA) for azimuthal changes around positions at 0 ° and 45 ° and elevation changes at 0 °, and Minimum Audible Distance (MAD) thresholds. Test protocol used a 2-down 1-up staircase method. Subjects included early (22 subjects) and late (13 subjects) blind children, sighted children (18 subjects), and a control group of sighted adults (17 subjects). Results for MAA@0 ° and MAD showed that blind children outperformed or performed comparably to sighted adults, both being better than sighted children. MAA@45 ° results were comparable for all groups. A reaching-sound-sources task showed comparable results in azimuth and elevation, but lower absolute and distance errors for blind children, underestimating ≈ 2.5 cm, with sighted subject underestimating ≈ 9.5 cm. Similarly, Voss et al. (2004) examined MAA and MAD thresholds for frontal and eccentric source positions using a same/difference task for early blind, late blind, and sighted control subjects. Results showed that early blind individuals had lower MAA thresholds for lateral positions, and that blind subjects in general had improved overall MAD and MAA for rear positions. Considering frontal horizontal sources, Dufour & Gérard (2000) showed that improved auditory localization performance extends to near-sighted individuals as well.

With improved techniques and understanding of localization task performance, some studies have questioned the actual protocol of a pointing task between sighted and blind subjects. Lewald (2002a) found poorer performance in vertical localization but raised the issue of the pointing task and the definition of the point of reference to extrapolate the indicated position. Zwiers et al. (2001b) went further, finding equal performance when taking into account a modified point of reference between sighted and blind subjects, being head or shoulder based. The need to examine more closely perceptual or cognitive differences between sighted and blind individuals has highlighted a variety of insights. Lewald (2002b) examined localization errors in the context of eccentric head positions. While errors for all groups were of similar magnitude, sighted subjects tended to undershoot, while blind subjects overshoot source position estimations. As a result, the following conclusion was offered: “…in contrast to the widespread opinion of compensation of visual loss by a general sharpening of audition, compensatory plasticity in the blind may specifically be related to enhanced processing of proprioceptive and vestibular information with the auditory spatial input." An equally interesting result was found by Ohuchi et al. (2006) in testing angular and distance localization for azimuthally located sources with and without head movement. Overall, blind subjects outperformed sighted control for all positions. For distance estimations, in addition to being more accurate, errors by blind subjects tended to be overestimations, while sighted control subject errors were underestimations, in accordance with numerous other studies. These studies indicate that one must take a second look at many of the accepted conclusions of auditory perception, especially spatial auditory perception, when considering the blind, who do not necessarily have the same error typologies due to different learning sensory conditions. A number of studies, such as Weeks et al. (2000), have focused on neural plasticity, or changes in brain functioning, evaluated for auditory tasks between blind and sighted subjects. Results by both Elbert et al. (2002) and Poirier et al. (2006) have shown increased activity in typically visual areas of the brain for blind subjects.

While localization, spectral analysis, and other basic tasks are of significant importance in understanding basic auditory perception and differences that may exist in performance ability between sighted and blind individuals, these performance differences are inherently limited by the capacity of the auditory system. Rather, it is in the *exploitation* of this acoustic and auditory information, requiring higher level cognitive processing, where blind individuals are able to excel relative to the sighted population. Navigational tasks are one instance where this seems to be clear. Strelow & Brabyn (1982) performed an experiment where subjects were to walk a constant distance from a simple straight barrier, being a wall or series of poles at 2 m intervals (diameter 15 cm or 5 cm), without any physical contact to the barrier. Footfall noise and finger snaps were the only information. With 8 blind and 14 blindfolded sighted control subjects, blind subjects clearly outperformed sighed subjects, some of whom claimed the task to be impossible. The results showed that blindfolded subjects performed overall as well in the wall condition as blind subject in the two pole conditions. Morrongiello et al. (1995) tested spatial navigation with blind and sighted children (ages 4.5 to 9 years). Within a carpeted room (3.7 m × 4.9 m), four tactile landmarks were placed at the center of each wall. Subjects, blind or blindfolded, were guided around the room to the different landmarks in order to build a spatial cognitive map. The same paths were used for all subjects, and not all connecting paths were presented. This learning stage was performed with or without an auditory landmark condition, a single metronome placed at the starting position. Subjects were then asked to move from a given landmark to another, with both known and novel paths being tested. Different trajectory parameters were evaluated. Results for sighted subjects indicated improvements with age and with the presence of the auditory landmark. Considering only the novel paths, all groups benefited from the auditory landmark. Analyzing the final distance error, sighted children outperformed blind in both conditions with blind subjects in the auditory landmark condition performing comparably to blindfolded subjects without auditory landmark. It is noted that due to the protocol used, it was not possible to separate auditory landmark and learning effect.

3. Virtual interactive environments for the blind: Academic context

Substantial amounts of work attest to the capacity of the blind and visually impaired to navigate in complex environments without relying on visual inputs (e.g., Byrne & Salter (1983); Loomis et al. (1993); Millar (1994); Tinti et al. (2006)). A typical experiment consists of having blind participants learn a new environment by walking around it, with guidance from the experimenter. How the participants perform mental operations on their internal representations of the environment is then assessed. For example, participants are invited to estimate distances and directions from one location to another (Byrne & Salter (1983)). Results from these experiments seem to attest that blind individuals perform better in terms of directional and distance estimation if the location of the experiment is familiar (e.g. at home) rather than unfamiliar.

Beyond the intrinsic value of the outputs of the research programs reported here, more information still needs to be collected on the conditions in which blind people use the acoustic information available to them in an environment to build a consistent, valid representation of it. 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. Is it the case that a learning procedure based upon the systematic exploitation of acoustic cues prepares a visually impaired person to move safely in a new and intricate environment? It then needs to be noted that blind people, who have to learn a new environment in which they will have to navigate, use typically special procedures. For instance, when a blind person gets a new job in a new company, it is common for him/her to begin by visiting the building late in the evening: the objective is to acquire some knowledge of the spatial configuration and of the basic features of the acoustical environment (including reverberation effects, sound of their steps on various floor surfaces, etc.). Later on, the person will get acquainted with the daily sounds attached to every part of the environment.

The following sections present a series of three studies which have been undertaken in order to better understand behaviours in non-visual complex auditory environments where spatial cognition plays a major role. A variety of virtual auditory environments and experimental platforms have been developed and put to the service of cognitive science studies in this domain, with special attention to issues with the visually impaired. These studies help both in improving the understanding of spatial cognitive processing as well as highlighting the current possibilities and limitations of different 3D audio technologies in providing sufficient spatial auditory information to subjects.

The first study employs a full-scale immersive virtual audio environment for the investigation of spatial cognition and localisation. Similar in concept to Morrongiello et al. (1995), this study provides for a more complex scene, and more complex interactions for study. As not all experiments can be performed using a full-scale immersive environment, the second study investigates the need for head-tracking by proposing a novel blind active virtual exploration task. The third and final study investigates spatial cognition through architectural exploration by comparing spatial and architectural understanding in real and virtual environments by blind individuals.

4. Study I. Mental imagery and the acquisition of spatial knowledge without vision: A study of blind and sighted people in an immersive audio virtual environment

Visual imagery can be defined as the representation of perceptual information in the absence of visual input (Kaski (2002)). In order to assess whether visual experience is a pre-requisite for image formation, many studies have focused on the analysis of visual imagery in congenitally blind participants. However, only few studies have described how visual experience affects the metric properties of the mental representations of space (Kaski (2002); Denis & Zimmer (1992)).

This section presents a study that was the product of a joint effort of different research groups in different areas for the investigation of a cognitive issue through the development and implementation of a general purpose Virtual Reality (VR) or Virtual Auditory Display (VAD) environment. The aim of this research project was the investigation of 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 (congenitally or early blind, 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. Details of this study, the system architecture and the analysis of the results, can be found in Afonso et al. (2005a);Afonso et al. (2005b);Afonso et al. (2005c);Afonso et al. (2010).

4.1 Mental imagery task using a tactile/haptic scene (background experiment)

The development of the VAD experiment followed the results of an initial study performed concerning the evaluation of mental imagery using a tactile or haptic interface. Six imaginary objects were located on the perimeter of a physical disk (diameter 50 cm) placed upright in front of the participants. The locations of these objects were learned by the participants exploiting two different modalities. The first one was a verbal description of the configuration itself, while the second one involved the experimenter placing the hand of the participant at the appropriate positions. After having acquired knowledge about the configuration of the objects through one of the two modalities, the participants were asked to create a mental representation of a given spatial configuration, and then to compare distances between the objects situated on the virtual disk.

The results showed that independent of the type of visual deprivation experienced by the participants and of the learning modality, all participants were able to create a mental representation of the configuration that preserved the metric relations between the objects. The precision of the spatial cognitive maps was evaluated using a mental scanning paradigm. The task consisted in mentally imagining a point moving between two objects, subjects responding when the trajectory was completed. A correlation between response times and scanned distances was obtained for all experimental groups and for both modalities. It was noted that blind subjects needed more time than sighted in order to achieve the same level of performance for all conditions.

The examined hypothesis was that congenital blind individuals, who are not expected to generate visual mental images, are nevertheless proficient at using mental simulation of trajectories. Sighted individuals would be expected to perform better, having experience in generating visual mental images. While no difference was found in precision, a significant difference was found in terms of response times between blind and sighted participants. A new hypothesis attempts to explain this difference by examining the details of the task (allocentric vs. egocentric) as being the cause, and not other factors. This hypothesis could explain the difference in the processing times needed by blind people in contrast to the sighted, and could explicate the tendency for the response times of blind individuals to be shorter after the haptic exploration of the configuration.

In order to test this hypothesis, a new experimental system was designed in which the task was conceived to be more natural for, even to the advantage of, blind individuals. An egocentric spatial scene, rather than the allocentric scene used in the previously described haptic task, was used. An auditory scene was also chosen.

4.2 An immersive audio interface

A large-scale immersive VAD environment was created in which participants could explore and interact with virtual sound objects located within an environment.

The scene in which the experiment took place consisted of a room (both physical and virtual) in which six virtual sound objects were located. The same spatial layout configuration and test positions were employed as in the previous haptic experiment. Six "domestic" ecological sound recordings were chosen and assigned to the numbered virtual sound sources: (1) running water, (2) telephone ringing, (3) dripping faucet, (4) coffee machine, (5) ticking clock, and (6) washing machine.

A virtual scene was constructed to match the actual experimental room dimensions. Monitoring the experiment by the experimenter was possible through different visual renderings of the virtual scene. The arrangement of the scene consisted of six objects representing the six sound sources located on the perimeter of a circle. A schematic view of the real and simulated environment and of the positions of the six sound sources is shown in Fig. 1. Participants were equipped with a head-tracker device, mounted on a pair of stereophonic headphones, as well as with a handheld tracked pointing device, both of which were also included in the scene graph. Collision detection was employed to monitor if a participant approached the boundaries of the physical room or the limits of the tracking system in order to avoid any physical contact with the walls during the experiment. A spatialized auditory alert, wind noise, was then used to warn the participants of the location of the wall in order to avoid contact.

The balance between direct and reverberant sound energy is useful in the perception of source distance (Kahle (1995)). It has also been observed that the reverberant energy, and especially a diffuse reverberant field, can negatively affect source localization. As this study was primarily concerned with a spatially precise rendering, rather than a realistic room acoustic experience, the reverberant energy was somewhat limited. Omitting the room effect creates an "anechoic" environment, which is not habitual for most people. To create a more realistic environment for which the room effect was included, an artificial reverberation was used with a reverberation time of 2 s. To counteract the negative effect on source localization, the direct to reverberant ratio was defined as 10 dB at 1 m. The design goal was for distance perception and precision localisation to be achieved through dynamic cues and subject displacements.

The audio scene was rendered over headphones using binaural synthesis (Begault (1994)) developed in the *MaxMSP*[1] environment. A modified version of *IRCAM Spat*[2] was also developed which allowed for the individualization of Inter-aural Time Delay (ITD) based on head circumference, independent of the selected Head Related Transfer Function (HRTF). The position and head orientation of the participant was acquired using a six Degrees-of-Freedom (6DoF) electromagnetic tracking system. Continuously integrating the updated external positional information, the relative positions of the sound sources were calculated, and the sound scene was updated and rendered, ensuring a stable sound scene irrespective of subject movements. The height of the sound sources was normalized relative to the subject's head height (15 cm above) in order to avoid excessive sound pressure levels when sources were approached very closely. An example of the experiment showing the different phases, including the subjective point of view binaural audio rendering, can be found on-line[3].

---

[1] http://www.cycling74.com

[2] http://forumnet.ircam.fr/692.html

[3] http://www.limsi.fr/Rapports/RS2005/chm/ps/ps11/ExcerptExpeVR.mov

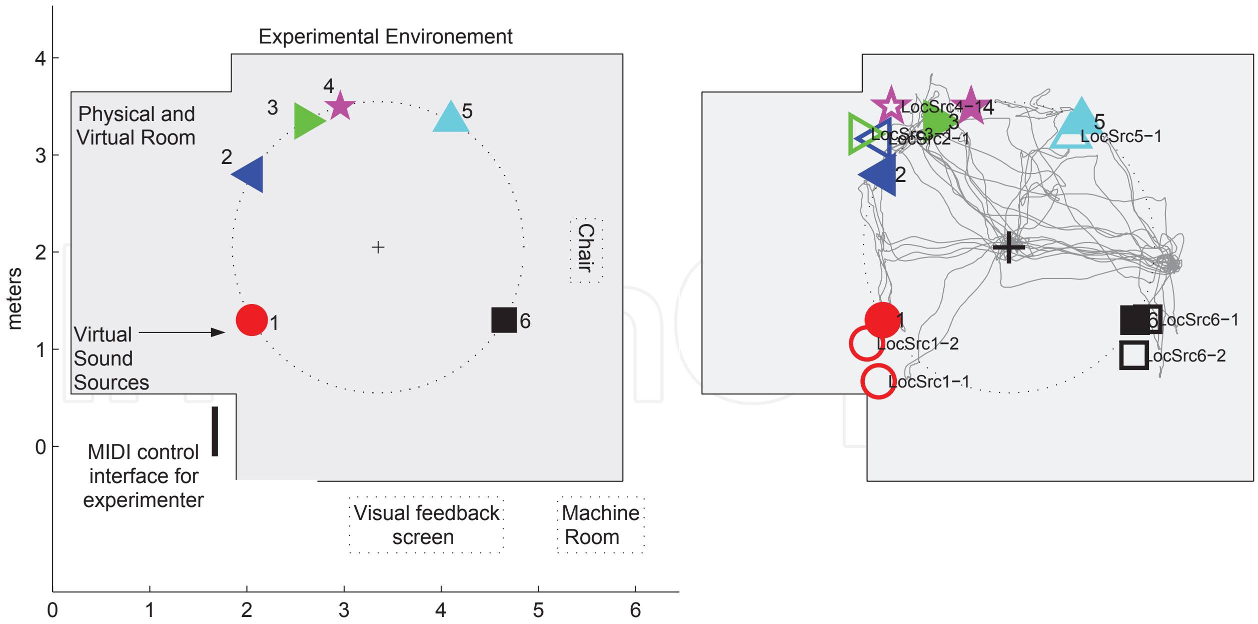


Fig. 1. Schematic view (left) of the real and simulated environment, together with the six sound sources and the reference point chair. Sample visualization (right) of experimental log showing participant trajectory and repositioned source locations (labelled LocSrc*n*-*pass*).

4.3 The task

A total of 54 participants took part in this study. Each one belonged to one of three groups: congenitally or early blind, late blind, and blindfolded sighted. An equal distribution was achieved between the participants of the three groups according to gender, age, and educational and socio-cultural background. These groups were split according to two learning conditions (see Section 4.3.1). Each final group comprised five women and four men, from 25 to 59 years of age.

4.3.1 Learning phase

The learning phase was carried out exploiting one of the two previously tested learning methods: Verbal Description (**VD**) and Active Exploration (**AE**). To begin, each participant was familiarised with the physical room and allowed to explore it for reassurance. They were then placed at the centre of the virtual circle (see Fig. 1) which they were informed had a radius of 1.5 m, and on which the six virtual sound sources were located.

For groups **VD**, the learning phase was passive and purely verbal. The participants were centred in the middle of the virtual circle and informed about the positions of the sound sources by first hearing the sound played in mono (non-spatialized), and then by receiving a verbal description, performed by the experimenter, about its location using conventional clock positions, as are used in aerial navigation, in clockwise order. No verbal descriptions of sound sources were ever used by the experimenter.

For groups **AE**, the learning phase consisted of an active exploration of the spatial configuration. Participants were positioned at the centre of the virtual circle. Upon continuous presentation of each sound source individually (correctly spatialized on the circle), participants had to physically move from the centre to the position of each sound source.

In order to verify that participants correctly learned the spatial configuration, each group was evauated. For groups **AE**, participants returned to the centre of the virtual circle where each sound source was played individually, non-spatialized (mono), in random order, and participants had to point (with the tracked pointer) to the location of the sound sources. The response was judged on the graphical display. The indicated position was valid if the pointer intersected with a sphere (radius = 0.25 m) on the circle (radius = 1.5 m), equating to an angular span of 20 ° centred on the exact position of the sonic object. For groups **VD**, participants had to express verbally where the correct source location was, in hour-coded terms. Errors for both groups were typically of the type linked to confusions between other sources rather than absolute position errors. In the case of any errors, the entire learning procedure was repeated until the responses were correct.

4.3.2 Experimental phase

Following the learning phase, each participant began the experiment standing at the centre of the virtual circle. One sound source was briefly presented, non-spatialized and randomly selected, whose correct position they had to identify. To do this, participants were instructed to place the hand-tracked pointer at the exact position in space where the sound object should be. The height component of the responses was not taken into account in this study. When participants confirmed their positional choice, the sound source was re-activated at the position indicated and remained active (audible) while each subsequent source was added. After positioning the first sound source, participants were led back to the reference chair (see Fig.1). All subsequent sources were presented from this position, rather than from the circle centre. This change of reference point was intentional in order to observe the different strategies used by participants to reconstruct the initial position of sound objects, such as directly walking to the source position or walking first to the circle centre. After placing the final source, all sources were active and the sound scene was complete. This was the first instance in the experiment when the participants could hear the entire scene.

Participants were then returned to the centre of the virtual circle, from where they were allowed to explore the completed scene by moving about the room. Following this, they were repositioned at the centre, with the scene still active. Each sound source was selected, in random order, and participants had the possibility to correct any position they judged incorrect using the same procedure as before.

4.4 Results

Visualization of the experimental phase is possible using the logged information, of which an example is presented in Fig. 1. One can see for several sources two selected positions, equating to the first pass position, and the second pass, refined position.

Evaluation of the experimental phase consisted in measuring the discrepancy between the original spatial configuration and the recreated sound scene. Influence of the learning modality on the preservation of the metric and topological properties of the memorized environment was analyzed in terms of angular, radial, and absolute distance errors as compared with the correct location of the corresponding object.

A summary of these errors is shown in Fig. 2. An ANalysis Of VAriance (ANOVA) was performed on the errors taking into account learning condition and visual condition for each group. Analysis of each error is discussed in the following sections.

4.4.1 Radial error

Radial error is defined as the radial distance error, calculated from the circle centre, between the position of the sound source and the actual position along the circle periphery. For both verbal learning and active exploration, participants generally underestimated the distances

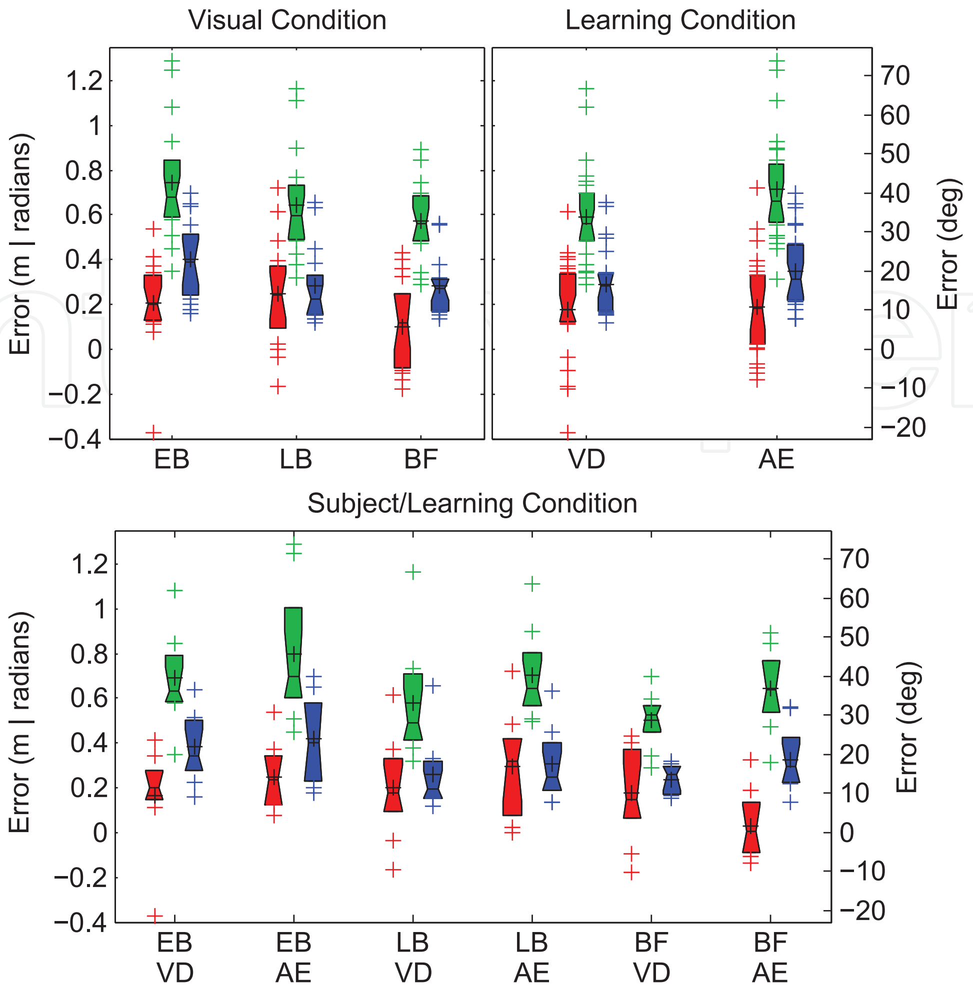


Fig. 2. Overview of the errors collapsed over visual condition (top left), learning condition (top right) and crossed effects (bottom). Radial errors (meters) in red, distance errors (meters) in green, and angular errors (radians left axis, degrees right axis) in blue. Learning conditions are Active Exploration, **AE**, and Verbal Description, **VD**. Visual conditions are Early Blind, **EB**, Late Blind, **LB**, and BlindFolded, **BF**. Black **+** indicate data mean values, notches indicate median values and confidence intervals, and coloured **+** indicate data outliers.

(a positive error) by the same amount (mean = 0.2 m), with similar standard deviation (0.3 m and 0.4 m, respectively). There was no difference among the three groups; each one underestimated the distance with a mean error of 0.2 m for congenitally blind (std = 0.3) and late blind (std = 0.4), and a mean error of 0.1 m for blindfolded (std = 0.3). Interestingly, a significant difference was found for blindfolded participants who learned the spatial configuration from a verbal description, underestimating radial positions (mean = 0.2 m, std = 0.3) when compared with an active exploration (mean = 0.0 m, std = 0.4) [F(2,48) = 3.32; p = 0.045].

4.4.2 Absolute distance error

Absolute distance error is defined as the distance between the original and selected source positions. Results show a significant effect of learning condition. Active exploration of the virtual environment resulted in better overall estimation of sound source positions (mean = 0.6 m, std = 0.3) as compared to the verbal description method (mean = 0.7 m, std = 0.4)

[F(1,48) = 4.29, p = 0.044]. The data do not reflect any significant difference as a function of visual condition (congenitally blind, mean = 0.7 m, std = 0.4; late blind, mean = 0.6 m, std = 0.3; blindfolded, mean = 0.6 m, std = 0.3).

4.4.3 Angular error

Angular error is defined as the absolute error in degrees, calculated from the position designated by participants in comparison to the circle centre of the reference position of the corresponding sound source. There was no significant difference between learning conditions: verbal description (mean = 17 °, std = 14 °) and active exploration (mean = 20 °, std = 17 °). Congenitally blind participants made significantly larger angular errors (mean = 23 ° , std = 17 °) than late blind (mean = 16 °, std = 15 °) [F(1,32) = 4.52; p = 0.041] and blindfolded sighted participants (mean = 16 °, std = 13 °) [F(1,32) = 6.08; p = 0.019].

4.5 Conclusion

The starting hypothesis was that the learning through active exploration would be an advantage to blind participants when compared to learning via verbal description. If true, this would confirm results of a prior set of experiments which showed a gain in performance of mental manipulations for blind people following this hypothesis (Afonso (2006)). A second hypothesis concerned sighted participants, who were expected to benefit more from a verbal description, being more adapt at generating a visual mental image of the scene, and thus being able to recreate the initial configuration of the scene in a more precise manner.

Considering the scene recreation task, these results suggest that active exploration of an environment enhances absolute positioning of sound sources when compared to verbal description learning. The same improvement appears with respect to radial distance errors, but only for blindfolded participants. Results show that participants underestimated the circle size, independent of the learning modality except for the case of blindfolded participants, with a mean position error close to zero, and that they clearly benefited from learning with perception-action coupling. These results are not in line with previous findings such as Ohuchi et al. (2006) in which blind subjects performed better at distance estimation for real sound sources using only head rotations and verbal position reporting. It clearly appears that an active exploration of the environment improves blindfolded participants' performance, both in terms of absolute position and size of the reconstructed configuration.

It has also been found that subjects blind from birth made significantly more angular positioning errors than late blind or blindfolded groups for both learning conditions. These data are in line with the results of previous studies involving spatial information processing in classic real (non virtual) environments (Loomis et al. (1998)).

5. Study II: A study on head tracking

This study focuses on the role of the Head Movements (HM) a listener uses in order to localize a sound source. Unconscious HM are important for resolving front-to-back ambiguities and for improving localization accuracy (see Wenzel (1998); Wightman & Kistler (1999); Minnaar et al. (2001)). However, previous studies regarding the importance of HM have all been carried out in static situations (participants at a fixed position without any positional displacement). The aim of this experiment is to investigate whether HM are important when individuals are allowed to navigate within the sound scene. In the context of future applications using VAD, it is useful to understand the importance of head-tracking. In this instance, a virtual environment was created employing a joystick for controlling displacement. Elements of this study have been presented by Blum et al. (2006), and additional details can also be found on-line[4].

5.1 Binaural rendering and head tracking

A well-known issue related to the use of non-tracked binaural technology consists in the fact that under normal headphone listening conditions, the sound scene follows HM, such that the scene remains defined in the head-centred reference frame, not in that of the external world, making it unstable relative to HM. In this situation, the individual is unable to benefit from binaural dynamic cues. However, with head orientation tracking, it is possible to update the sound scene relative to the head orientation in real time, correcting this artefact.

In the present experiment, two conditions have been tested: actual orientation head-tracking versus virtual head rotations controlled via joystick. Participants with head-tracking can have pertinent acoustic information from HM as in a natural 'real' situation, whereas participants without head-tracking have to extrapolate cues from other control movements. The hypothesis is that an active exploration task with linear displacements in the VAD is sufficient to resolve localization ambiguities, implying that tracking HM is not always necessary.

5.2 Experimental task

The experiment followed a 'game like' scenario of bomb disposal, and was carried out with sighted blindfolded subjects. Bombs (sound sources simulating a ticking countdown) were located in a virtual open space. Participants had to find them by navigating to their position, using a joystick (displacement control and virtual head rotation relative to the direction, of motion using the twist of the joystick) to move in the VAD. The scene was rendered over headphones (see Section 4.2 for a description of the binaural engine used). For the head-tracked condition, an electromagnetic tracker was employed with a refresh rate of 20 Hz. To provide a realistic auditory environment, artificial reverberation was employed. The size of the virtual scene, and the corresponding acoustics, was chosen to correspond to an actual physical room (the Espace de Projection, *Espro*, at IRCAM) with its variable acoustic characteristics in its more absorbing configuration (reverberation time of 0.4 s). Footstep sounds were included during movement, rendered to aid in the perception of displacement and according to the current velocity.

In the virtual environment, the relation between distances, velocity, and the corresponding acoustic properties was designed so as to fit a real situation. Forward/backward movements of the joystick allowed displacement respectively forward and backward in the VAD. The maximum speed, corresponding to the extreme position, was 5 km/h, which is about the natural human walking speed. With left/right movements, participants controlled body rotation angle, which relates to the direction of displacement. Translation and rotation could be combined with diagonal manipulations. The mapping of lateral joystick position, δx, to changes in navigation orientation angle, α, was based on the relation: $\alpha = (\delta x / x_{max}) 50^\circ \delta t$; where x_{max} is the value corresponding to the maximum lateral position of the joystick, and δt the time step between two updates of δx.[5] For the material used, this equation provides a linear relation between α and δx with a coefficient of 0.001.

The design of the task was centered on the principle that, as with unconscious HM, linear displacements and a stable source position would allow for the resolution of front-back

---

[4] http://rs2007.limsi.fr/index.php/PS:Page_16

[5] http://www.openscenegraph.org/

confusions. To concentrate on the unconscious aspect, a situation involving two concurrent sources was chosen. While the subject was searching for one bomb, the subsequent target would begin ticking. As such, the conscious effort was focussed on the current target, while the second target's position would become more stable in the mental representation of the scene. This was thought to incite HM for the participant for localizing the new sound while keeping a straight movement toward the current target. As two sources could be active at the same time, two different countdown sounds were used alternatively with equal normalized level.

Each test series included eight targets. The distance between two targets was always 5 m. In order to enforce the speed aspect of the task, a time limit (60 s) was imposed to reach each target (defuse the bomb), after which the bomb exploded. The subsequent target would begin ticking when the subject arrived within a distance of 2 m from the current target. In the event of a failed target, the participant was placed at the position of the failed target and would then resume the task towards the next target. Task completion times and success rates were used to evaluate the effects of the different conditions.

A target was considered found and defused when the participant arrived within a radius of 0.6 m. This 'hit detection radius' of 0.6 m corresponds to an angle of $\pm 6.8\,^\circ$ at a distance of 5 m from the source, which is the mean human localization blur in the horizontal plane (Blauert (1996)). As a consequence, if the participant oriented him/herself with this precision when starting to look for a target, this could be reached by going straightforward.

The experiment was composed of six identical trials involving displacement along a succession of eight segments (eight sources to find in each trial). The first trial was considered a training session, and the last segment of each trial was not taken into account as only a single target signal was present for the majority of the search.

In total, $5 \times 6 = 30$ segments per participant were analyzed. The azimuthal angles made by the six considered segments of each trial were balanced between right/left and back/front ($-135\,^\circ$, $-90\,^\circ$, $-45\,^\circ$, $45\,^\circ$, $90\,^\circ$, $135\,^\circ$). Finally, to control a possible sequence effect, two different segment orderings were created and randomly chosen for each participant.

5.3 Recorded data

Twenty participants without hearing deficiencies were selected for this study. Each subject was allocated to one of the two head-tracking conditions (with or without). An equal distribution was achieved between the participants of the two groups according to gender, age, and educational and socio-cultural background. Each group comprised five women and five men with a mean age of 34 years (from 22 to 55 years, std = 10).

Result analysis was based on the following information: *hit time* (time to reach target for each segment), *close time* (time to get within 2 m from target, when the subsequent target sound starts), and the *total percentage* of successful hits (bombs defused).

Position and orientation of the participant in the VAD were recorded during the entire experiment, allowing for subsequent analysis of trajectories. At the end of the experiment, participants were asked to draw the trajectory and source positions on a sheet of paper (the starting point and first target were already represented in order to normalize the adopted scale and drawing orientation).

5.4 Results

Large individual differences in hit time performance ($p < 10^5$) were observed. Some participants showed a mean hit time more than twice the quickest ones. Percentage of successful hits varied from 13% to 100%, and the participants that were quicker in completing the task, obtained a higher percentage of hits. In fact, some participants were practically unable to execute the task while others exhibited no difficulty. Performance measures of mean hit times and total percentage hit were globally correlated with a linear correlation coefficient of -0.67 ($p = 0.0013$).

The influence of the source position sequence (two different orderings were randomly proposed) and of the type of source (two different sounds were used) was tested. No effect was found for these two control variables.

Analysis of hit times and head-tracked condition did not reveal any significant effect. Mean hit times of the two groups were very similar (19.8 s versus 20.4 s). Table 1 shows that participants in the head-tracked condition for HM did not perform better than those in the non-tracked condition.

A significant effect was found for subject age. Four age groups were defined with four participants between 20 and 25 years, six between 25 and 30, six between 30 and 40 and four between 40 and 60. Table 1 shows the performances for each age group. Young participants had shorter hit times and higher percentage of hits if compared with older ones. A significant effect of age ($p < 0.0001$) and a significant gender×age interaction ($p = 0.0007$) were found: older women had more difficulty in executing the task.

	HM Tracking		Age Groups				Videogame Experience	
Condition	No	Yes	20-25	25-30	30-40	40-60	No	Yes
mean Hit Time (s)	19.8	20.4	17.0	20.4	19.7	29.4	22.1	19.0
Standard Deviation (s)	11.1	11.9	10.3	11.9	9.6	15.2	12.2	10.8
% Hit Sources	86%	80%	92%	94%	94%	35%	69%	94%

Table 1. Performance results as a function of tracking, age, and video game experience.

In a questionnaire filled in before the experiment, participants were asked to report whether they had previous experience with video games. Eleven participants reported they had such experience, while the remaining nine participants did not. Table 1 shows that the experienced group had higher performances results. There was a significant effect of this factor on hit times ($p = 0.004$), and the group with video game experience had 94% hits versus only 69% for the other group. Not surprisingly, individuals familiar with video games seemed more comfortable with immersion in the virtual environment and also with joystick manipulation. This can be related to the age group observation since no participant from group [40-60] reported any experience with video games.

A significant learning effect was found ($p = 0.0047$) between trials, as shown in Table 2. This effect was most likely due to a learning effect of navigation within the VAD rather than a memorization of the position of the sources, since participants did not experience any sequence repetition and reported that they treated each target individually. Results of the post navigation trajectory reconstruction task confirm this by the fact that participants were unable to precisely draw the path of a trial on a sheet of paper when they were asked to do so at the end of the experiment. This lack of reconstruction ability is in contrast to the previous experiment (see Section 4), where subjects were able to reconstruct the sound scene after physical navigation. This can be seen as an argument in favour of the importance of the memorization of sensorimotor (locomotor) contingencies for the representation of space.

Through inspection of the different trajectory paths, it was observed that front/back confusions were present for participants in both tracking conditions. In Fig. 3A and Fig. 3B,

Learning effect					
Trial	2	3	4	5	6
Hit Time Mean (sec.)	22.1	22.4	20.2	19.0	17.0
Standard Deviation	12.5	12.6	11.5	9.8	9.9
% of hit sources	78%	82%	85%	81%	91%

Table 2. Performance as a function of trial sequence over all subjects.

two trajectories with such inversion are presented for two participants in the ‘head-tracking’ condition. Example A shows front/back confusion in the path between sources 3 and 4: the participant reaches source 3, source 4 is to the rear, but the subject moves forward in the opposite direction. After a certain distance is travelled, the localization inversion is realized and the subject correctly rotates again to go back in the correct direction. Fig. 3B shows a similar event between sources 1 and 2. Overall, in comparing the head orientation vector and the movement orientation vector, participants in the head-tracked condition did not appear to use HM to resolve localization ambiguities, focusing on the use of the joystick, keeping the head straight and concentrating only on frontal targets. It is apparent that rotations were typically made with the joystick at each source to decide the correct direction.

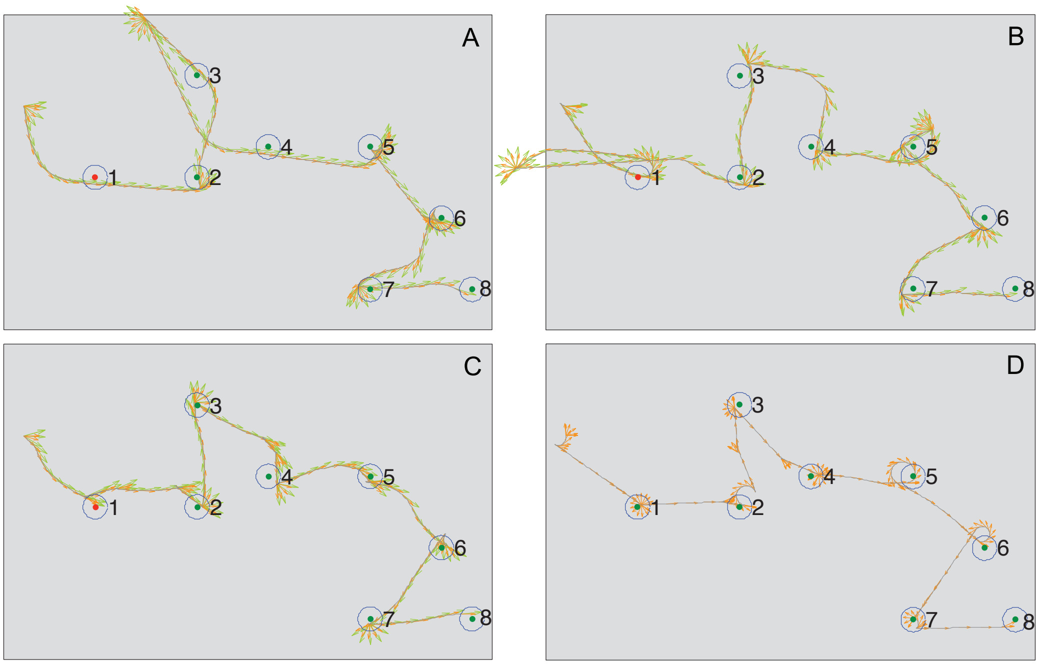


Fig. 3. Examples of trajectories of different participants with (ABC) and without (D) head-tracking. Arrows indicate movement orientation (orange) and the head orientation (green). A-B: examples of front/back confusion. C-D: typical navigation strategies with (C) and without (D) head-tracking condition.

5.4.1 Discussion and perspectives

The inclusion of head-tracking was not found to be necessary for the task proposed in this experiment. Movements of the joystick and virtual displacement were considered sufficient for the participants to succeed in the task. However, the use of a joystick elicits some questions pertaining to subject experience with video games and to the effect on task performance, as well as to the apparent lack of use of HM even when available.

Participants seem to have transferred vestibular modality toward the use of the joystick. This is supported by the typical navigation strategy observable in the participants' trajectories where rotations were made with the joystick (Fig. 3C and Fig. 3D). It is not yet clear how this finding can be extended to other tasks which require a more complex understand of the sound scene. As the subjects were not able to recount the positions of the different targets or their trajectories, it is possible that HM are still required for more complex spatially related tasks.

6. Study III. Creating a Virtual reality system for Visually impaired persons

This research results from collaboration between researchers in psychology and in 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. Focusing on the needs of the blind, creation of 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.

A comparison of three types of learning has been performed: *in situ* real displacement, passive playback of a recorded navigation (with and without HM tracking), and active navigation in a virtual architecture. For all conditions, only acoustic cues are employed.

6.1 Localisation versus spatial perception

Sound source localisation in an anechoic environment is a special and quite unnatural situation. It is more typical to hear sound sources with some amount of reflections, even in outdoor environments, or with a high density of reflections in reverberant spaces. These additional acoustic path returns from the same source can cause certain impairments, such as source localisation confusion and degradation of intelligibility. At the same time, these additional acoustic signals can provide information regarding the dimensions, material properties, as well as cues improving sound source localisation.

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 localisation of the other signals resulting from reflections on 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 at. (1949)), and it allows for the localisation of a sound source even in situations when the reflections of the sound are actually louder than the direct signal. There are of course situations where errors occur, if the reflected sound is sufficiently louder and later than the direct sound. Other situations can also be created where false localisation occurs, such as with the Franssen effect (Hartmann & Rakerd (1989)), but those are not the subject of this work. The later arriving signals, while not being useful for localization, are used to interpret the environment.

The ability to directionally analyse 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 Section 3, information about the spatial configuration of a given environment is mainly gathered though sight, and not through hearing. For this reason, a sighted individual will find information about the direction of the reflected signal components redundant, while a blind individual will need this information in order to gather knowledge about the spatial configuration of an environment. Elements in support of this will be given in Section 6.4 and 6.4.3, 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.

It is clear that most standard interactive VR systems (e.g. gaming applications) are visually-oriented. While some engines take into account source localisation of the direct sound, reverberation is most often simplified and the spatial aspects neglected. Basic reverberation algorithms are not designed to provide such geometric information. Room acoustic auralization systems though should provide such level of spatial detail (see Vorländer, (2008)). This study proposes to compare the late acoustic cues provided by a real architecture with those furnished both by recordings and by using a numerical room simulation, as interpreted by visual 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.

6.2 Architectural space

In contrast to the previous studies, this one focuses primarily on the understanding of an architectural space, and not of the sound sources in the space. As a typical example, this study focuses on several (four) corridor spaces in a laboratory building. These spaces are not exceptionally complicated, containing a various assortment of doors, side branches, ceiling material variations, stairwells, and static noise sources. An example of one of the spaces used in this study is shown in Fig. 4. In order to provide reference points for certain validations, some additional sound sources were added. These simulated sources were simple audio loops played back over positioned loudspeakers.

6.3 Comparison of real navigation to recorded walkthrough

Synthesized architectural environments, through the use of numerical modelling, are necessarily limited in their correspondence to a real environment. In contrast, it can be hypothesized that a spatially correct recording performed in an actual space should be able to capture and allow for the reproduction of the actual acoustic cues, without the need to necessarily define or prescribe said cues.

In order to verify this hypothesis, two exploration conditions were tested within the four experimental corridors: real navigation and recorded walkthrough playback. In order to take into account the possible importance of HM, two recording methods were compared. The first, binaural recording, employs a pair of tiny microphones placed at the entrance of the ear canals. This recording method captures the fine detail of the HRTF but is limited in that the head orientation is encoded within the recording. The second method, Ambisonic recording, employs a spatial 3-dimensional recording. This recording, upon playback, can be rotated and as such can take into variations in head orientation during playback.

For the real navigation condition, a blind individual was equipped with in-ear binaural microphones (open meatus in order not to obstruct natural hearing) in order to monitor and be able to analyse any acoustic events. The individual then advanced along the corridor from

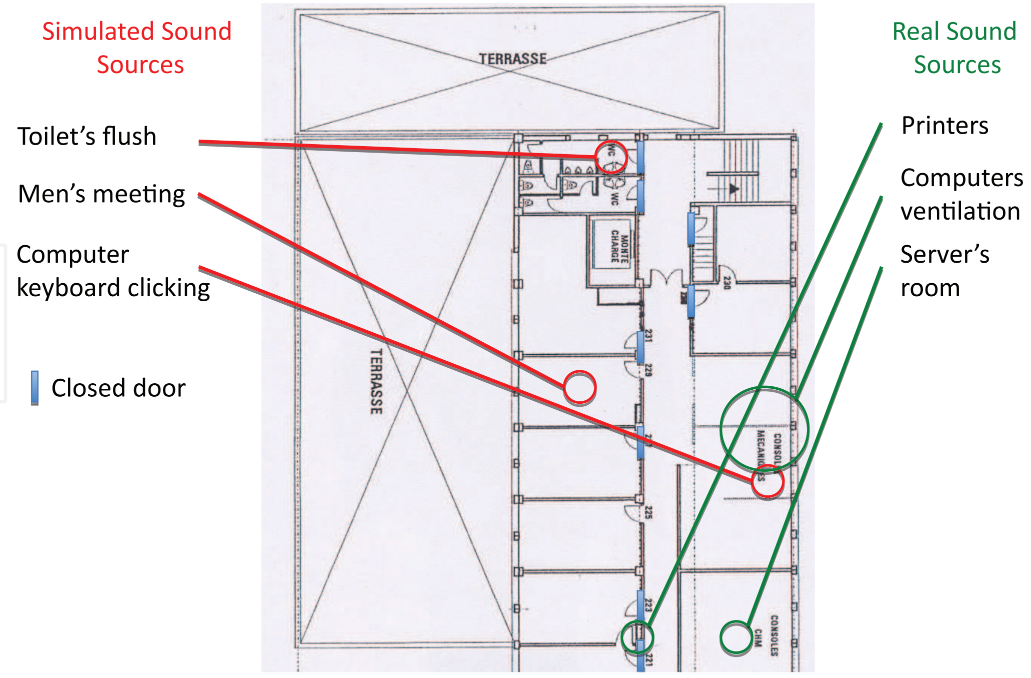


Fig. 4. Plan and positions of real and artificially simulated sound sources for environment 1.

one end to the other, and returned. No other navigation aides were used (cane, guide dog, etc.), but any movements or sounds were allowed. Contact with the environment was to be avoided, and the individual remarkably avoided any collisions. This navigation was tracked using a CCTV camcorder system with visual markers placed throughout the space for later calibration.

In order to have recordings for the playback conditions, an operator equipped with both binaural (in-ear *DPA 4060*) and B-Format (Gerzon (1972)) (*Soundfield ST250*) recording systems precisely repeated the path of the real navigation condition above. Efforts were made to maintain the same speed, and head movements, as well as any self-generated noises. This process was repeated for the four different environments.

6.3.1 Playback rendering system

In the Ambisonic playback condition the B-Format recording was then rendered over binaural headphones employing the approach of *virtual speakers*. This conversion from Ambisonic to stereo binaural signal was realized through the development and implementation of a customized software platform using *MaxMSP* and a head orientation tracking device (*XSens MTi*). The 3D sound-field recorded (B-Format signal) was modified in real-time performing rotations in the Ambisonics domain as a function of participant's head orientation. The rotated signal was then decoded on a virtual loudspeakers system with the sources placed on the vertices of a dodecahedron, at 1 m distance around the centre. These twelve decoded signals were then rendered as individual binaural sources via twelve instances of a binaural spatialization algorithm, which converts a monophonic signal to a stereophonic binaural signal (Fig. 5). The twelve binauralized virtual loudspeaker signals are then summed and rendered to the subject.

The binaural spatialization algorithm used was based on the convolution between the signal to be spatialized and a HRIR (*Head Related Impulse Response*) extracted from the Listen IRCAM database[6]. More information about this approach can be found in McKeag & McGrath. (1996). Full-phase HRIR were employed, rather than minimum-phase simplifications, in order to maintain the highest level of spatial information. A customization of the Interaural Time Differences (ITD), given the head circumference of the tested participant, and an HRTF selection phase were also performed as mentioned in the previously cited studies, so that an optimal binaural conversion could be performed.

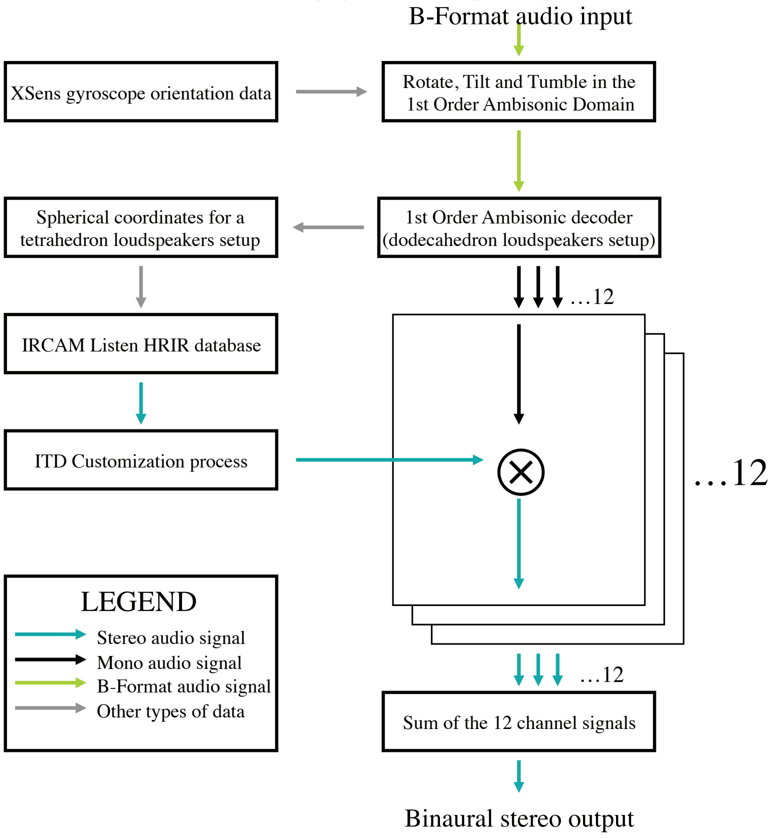


Fig. 5. Schematic representation of the Ambisonic to binaural conversion algorithm.

6.3.2 Protocol and Results: Real versus recorded walkthrough

Two congenitally blind and three late blind participants (two female, three male) took part in this experiment. Each subject was presented with one of the two types of recordings for two of the four environments. Participants were seated during playback.

The learning phase consisted of repeated listings to the playback until the participant felt they understood the environment. When presented with binaural renderings, participants were totally passive, having to remain still. Head orientation in the scene was dictated by the state of the recording. When presented with Ambisonic renderings, they had the possibility to freely perform head rotations, which resulted in real-time modification of the 3D sound environment, ensuring stability of the scene in the world reference frame. Participants were allowed to listen to each recording as many times as desired. As these were playback recordings, performed at a given walking speed, it was not possible to dynamically change the navigation speed or direction. Nothing was asked of the participants in this phase

---

[6] http://recherche.ircam.fr/equipes/salles/listen/

Two tasks followed the learning phase. Upon a final replay of the playback, participants were invited to provide a verbal description of every sound source or architectural element detected along the path. Following that, participants were invited to reconstruct the spatial structure of the environment using a set of LEGO® blocks. This reconstruction was expected to provide a valid reflection of their mental representation of the environment.

A similar task was demanded to one congenitally blind individual who performed a real navigation within the environments, and was used as a reference.

The verbal descriptions revealed a rather poor understanding of the navigated environments, which was confirmed by the reconstructions. Fig. 7 shows a map of one actual environment and LEGO® reconstruction for different participant conditions. For the real navigation condition, the overall structure and a number of details are correctly represented. The reconstruction shown for the binaural playback condition reflects strong distortions as well as misinterpretations, as assessed by the verbal description. The reconstruction shown following for the Ambisonic playback condition reflects similar poor and misleading mental representation.

Due to the very poor results for this test, indicating the difficulty of the task, the experiment was stopped before all participants completed the exercise. Overall, results showed that listening to passive binaural playback or Ambisonic playback with interactive HM did not allow blind people to build a veridical mental representation of the virtually navigated environment. Participants' comments about the binaural recordings pointed to the difficulties related to the absence of information about displacement and head orientation. Ambisonic playback, while offering head-rotation correction, still resulted in poor performance, worse in some cases relative to binaural recordings, because of the poorer localization accuracy provided by this particular recording technique. Neither condition was capable of providing useful or correct information about displacement in the scene. The most interesting result was that none of the participants understood that recordings were made in a straight corridor with openings on the two sides.

As a final control experiment, after the completion of the reconstruction task, participants were invited to actually explore one of the corridors. They confirmed that they could perceive exactly what they heard during playback, but that it was the sense of *their own displacement* that made them able to describe correctly the structure of the navigated environment. This corroborates findings of previous studies for which the gathering of spatial information is significant for blind individuals when learnt with their own displacements (see Section 4). Further analysis of the reconstruction task can be found in Section 6.4.1.

6.4 Comparison of real and virtual navigation

The results of the preliminary phase of the project outlined 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. The missing element seemed to be found in the lack of interactivity and free movement within the simulated environment. For this reason, a second experiment was developed, with the objective of delivering information about the spatial configuration of a closed environment and the positions of sound sources within the environment itself, exploiting interactive virtual acoustic models.

Two of the four closed environments from the initial experience were retained, for which 3D architectural acoustic models were created using the CATT-Acoustics software[7]. Within each

---

[7] http://www.catt.se

of these acoustic models, in addition to the architectural elements, the different sound sources from the real situation (both real and artificial) were included in order to be able to carry out a distance comparison task (see Section 6.4.1). A third, more geometrically simple model was created for a training phase in order for subjects to become familiar with the interface and protocol. The geometrical model of one experimental space is shown in Fig. 6.

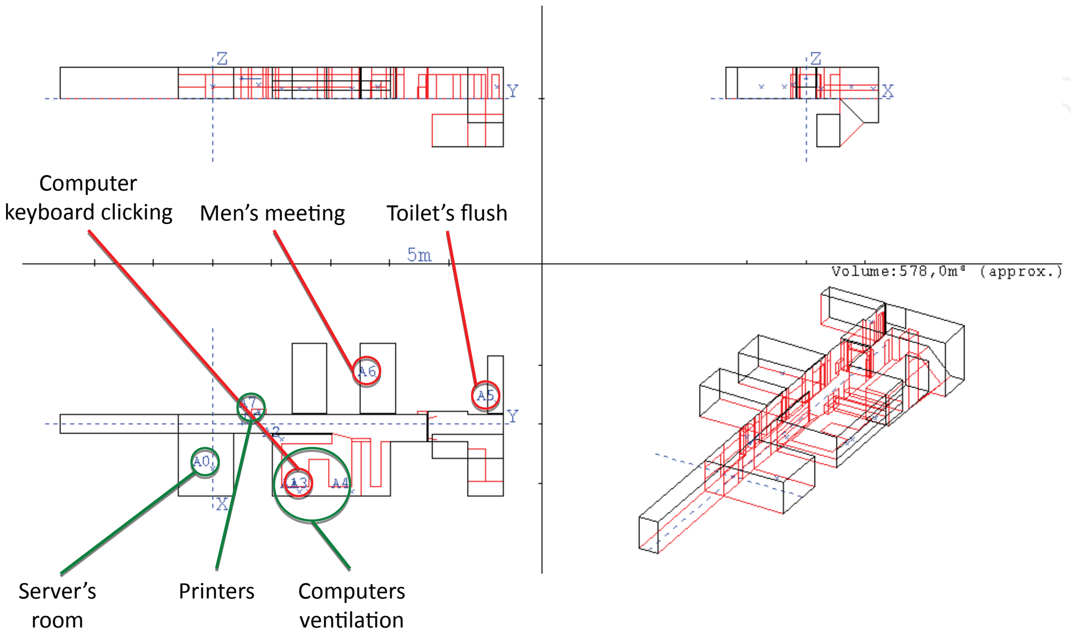


Fig. 6. Geometrical acoustic model of the first space including positions of real (green) and artificially simulated (red) sources.

After observations in the real navigation stage that blind individuals made extensive use of self-produced noises, such as finger snaps and footsteps, in order to determine the position of an object (wall, door, table, etc.) by listening to the reflections of the acoustic signals (see also Section 6.1), a simulation of these noises was included. With the various elements taken into account, a large number of spatial impulse responses were required for the virtual active navigation rendering. A 2^{nd} order Ambisonic rendering engine was used (as opposed to the prerecorded walkthough which used 1^{st} order Ambisonic) to improve spatial precision while still allowing for dynamic head rotation.

Due to the large number of concurrent sources and to the size of 2^{nd} order impulse responses (IR), a real-time accurate rendering was not feasible. Therefore, another approach was elaborated. As a first step, navigation was limited to one dimension only. Due to the fact that both environments were corridors, the user was given the possibility to move along the centreline. Receiver positions were defined at equally spaced positions along this line, at head height, as well as source positions at ground level (for footfall noise) and waist height (finger snap noise). In order to provide real-time navigation of such complicated simulated environments, it was decided to pre-calculate the 2^{nd} order Ambisonic signals for each position of the listener, and then to pan between the different signals during the real-time navigation, rather than performing all the convolutions in real-time, converting finally the Ambisonic signals to binaural using the same approach described in Section 6.3, modified to account for 2^{nd} order Ambisonic.

In the experimental condition, participants were provided with a joystick as a navigation device and a pair of headphones equipped with the head-tracking device (as in Section 6.3).

The footfall noise was automatically rendered in accordance with displacements in the virtual environment. The mobile self-generated finger snap was played each time the listener pressed a button on the joystick.

6.4.1 Protocol: Real versus virtual navigation

The experiment consisted in comparing two modes of navigation along two different corridors, with the possibility offered to the participants to go back and forth along the path at their will. Along the corridor, a number of sources were placed at specific locations, corresponding to those in the real navigation condition. In the real condition, two congenitally blind and three late blind individuals (three females, two males) participated for two corridors. In the virtual condition, three congenitally blind and two late blind individuals (three females, two males) explored the same two corridors.

The assessment of the spatial knowledge acquired in the two learning conditions involved two evaluations, namely a reconstruction of the environment using LEGO® blocks (as in Section 6.3.2) and a test concerning the mental comparison of distances. For the first navigated corridor, the two tasks were executed in one order (block reconstruction followed by distance comparison), while for the second learned corridor the order was reversed.

6.4.2 Block reconstruction

Several measures were made on the resulting block reconstructions: number of sound sources mentioned, number of open doors and staircases identified, number of perceived changes of the nature of the ground, etc. Beyond some distinctive characteristics of the different reconstructions (e.g. representation of wide or narrower corridor), no particular differences could be found between real and virtual navigation conditions; both were remarkably accurate as regards the relative positions of the sound sources (see example in Fig. 7). Door openings into rooms containing a sound source were well identified, while more difficulty was found for openings with no sound source present. Participants were also capable of distinctively perceiving the various surface material changes along the corridors.

An objective evaluation on how similar the different reconstructions are from the actual map of the navigated environment was carried out using bidimensional regression analysis (Nakaya (1997)). After some normalisation, the positions of the numerous reference points, both architectural elements and sound sources (93 coordinates in total) were compared with the corresponding point in the reconstructions, with a mean number of points of 46 ± 12 over all subjects. The bidimensional regression analysis results in a correlation index between the true map and the reconstructed map. Table 3 shows the correlation values of the different reconstructions for real and virtual navigation conditions, together with the correlations for the limited reconstructions done after the binaural and Ambisonic playback conditions, for the first tested environment. Results for the real and virtual navigation conditions are comparable, and both are greater than those of the limited playback conditions. This confirms the fact that playing back 3D audio signals, with and without head-tracking facilities, is not sufficient in order to allow the creation of a mental representation of a given environment due mainly to the lack of displacement information. On the other hand, via real and virtual navigation this displacement information is present, and the amelioration of the quality of the mental reconstruction is confirmed by the similar values in terms of map correlation. Furthermore, correlation values corresponding to the virtual navigation are slightly higher than those for real navigation, confirming the accuracy of the mental reconstruction in the first condition compared with the second.

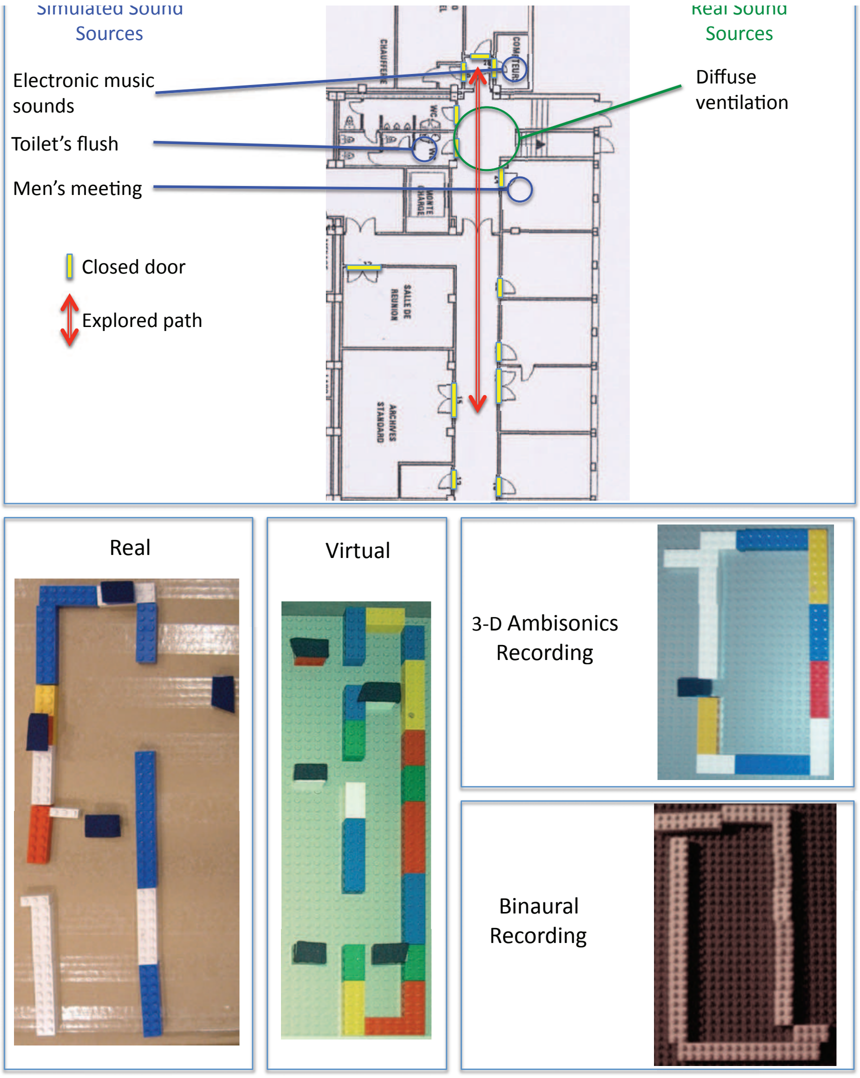


Fig. 7. Examples of LEGO® reconstructions following real navigation, virtual navigation, and binaural and Ambisonic playback.

Correlation of the LEGO® reconstruction				
Condition	Real	Virtual	Ambisonic Rec	Binaural Rec
Correlation Index mean	0.81	0.83	0.71	0.72
Standard Deviation	0.04	0.15	-	-

Table 3. Correlation and standard deviation for bidimensional regression analysis of reconstructions for architectural environment 1. (Std is not available for playback conditions as they contain only 1 entry each.)

6.4.3 Distance comparison

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 the longer, the frequency of correct responses is lower and the latency of responses is longer 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); Nordzij & Postma (2005)).

In addition to the starting and the arrival points, three sound sources existed along each path within the two navigated environments (1st: keyboard, men's voices and toilet flush; 2nd: women's voices, electronic sound, and toilet flush). All distances pairs, having one common item for each path (e.g., keyboard-men's voices / keyboard-toilet), have been considered. Distances were classified into three categories: small, medium, and large. Participants were presented with each pair of distances orally, and had to then indicate which was the longer of the two.

Analysis of the results focused on the frequency of correct responses. Table 4 shows the frequency of correct responses for the participants for both real and virtual navigation conditions.

Distance comparison						
Environment	Real			Virtual		
Distance type	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 4. Percent of correct responses for distance comparisons as a function of navigation condition.

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. Analysis is difficult as, for both conditions, results are near perfect for medium distances and perfect for large distances. The similarity of results for the two conditions is notable. 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.

Some interesting points were reported by the participants in the virtual navigation condition. They reported that for sound sources that were located at the left or at the right of the corridor, they perceived both the direct signal coming from the sound source and the reflected signal coming from the opposite direction (reflection off the wall), making it possible to locate both the source on one side and the reflecting object (in this case a wall) on the other. The finger snap sound (auditory feedback) was considered extremely useful for understanding some spatial configurations. Both these factors can be considered as extremely important results in light of what has been described in Section 6.1, corroborating the hypothesis that the developed application could indeed offer a realistic and well defined acoustical virtual reality simulation of a given environment, precise enough so that information about the spatial configuration of the total environment, not just source positions, can be gathered by visually impaired users solely through auditory exploration.

7. Acknowledgements

Studies were supported in part by a grant from the European Union (STREP Wayfinding, Contract 12959) and internal research grants from the LIMSI-CNRS (Action Initiative). Experiments conducted were approved by the Ethics Committee of the National Centre for

Scientific Research (*Comité Opérationnel pour l'Ethique en Sciences de la Vie*). The authors are grateful to Michel Denis, for supervision of the cognitive aspects of these studies, and to Christian Jacquemin, for his invaluable contributions to the design of the virtual environments. Finally, the authors would like to thank Amandine Afonso and Alan Blum, collaborators and co-authors of the various studies presented.

8. References

Afonso, A., Katz, B.F.G., Blum, A. & Denis, M. (2005a). Spatial knowledge without vision in an auditory VR environment, *Proc. of the 14th Meeting of the European Society for Cognitive Psychology*, August 31 - September 3, Leiden, the Netherlands.

Afonso, A., Katz, B.F.G., Blum, A., Jacquemin, C. & Denis, M. (2005b). 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.

Afonso, A., Katz, B.F.G., Blum, A. & Denis, M. (2005c). Mental imagery and the acquisition of spatial knowledge without vision: A study of blind and sighted people in an immersive audio virtual environment, *Proc. of the 10th European Workshop on Imagery and Cognition*, 28-30 June, St Andrews, Scotland.

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, No. 1, pp. 591-604.

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*, Vol. 27, pp. 105-122.

Begault, D. R. (1994). *3-D sound for virtual reality and multimedia*, Cambridge, MA: Academic Press.

Blauert, J. (1996). *Spatial Hearing, the Psychophysic of Human Sound Localization*, Cambridge, Massachusetts, USA: The MIT Press Cambridge.

Blum, A., Denis, M., Katz, B.F.G. (2006). Navigation in the absence of vision : How to find one's way in a 3D audio virtual environment?, *Proc. of the International Conference on Spatial Cognition*, September 12-15, Rome & Perugia, Italy.

Byrne, R. W., & Salter, E. (1983). Distances and directions in the cognitive maps of the blind. *Canadian Journal of Psychology*, Vol. 37, pp. 293-299.

Denis, M. & Zimmer, H. D. (1992). Analog properties of cognitive maps constructed from verbal descriptions, *Psychological Research*, Vol. 54, pp. 286-298.

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. (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*, Vol. 127, pp. 197-210.

Doucet, M.E., Guillemot, J.P., Lassonde, M., GagnÃl'é J.P., Leclerc, C., & Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals, *Experimental Brain Research*, Vol. 160, No. 2, pp. 194-202.

Dufour, A. & Gérard, Y. (2000). Improved auditory spatial sensitivity in nearsighted subjects, *Cognitive Brain Research*, Vol. 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*, Vol. 22, pp. 9941-9944.

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.

Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R.J., & Belin, P. (2004). Neuropsychology: Pitch discrimination in the early blind, *Nature*, Vol. 430, p. 309.

Hartmann, W.M. & Rakerd, B. (1989). Localization of sound in rooms IV: The Franssen effect, *J. Acoust. Soc. Am.*, Vol. 86, pp. 1366-1373.

Kahle, E. (1995) *Validation d'un modéle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d'opéras.* Doctoral dissertation, Université du Maine, Le Mans.

Kasky, D. (2002). Revision: Is visual perception a requisite for visual imagery?, *Perception*, Vol. 31, pp. 717-731.

Lessard, N., Paree, Lepore, F. & Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects, *Nature*, Vol. 395, pp. 278-280.

Lewald, J. (2002a). Vertical sound localization in blind humans, *Neuropsychologia*, Vol. 40, No. 12, pp. 1868 âĂŞ 1872.

Lewald J. (2002b). Opposing effects of head position on sound localization in blind and sighted human subjects, *European Journal of Neuroscience*, Vol. 15, pp. 1219-1224.

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*, Vol. 122, pp. 73-91.

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.

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.

Minnaar, P., Olesen, S.K., Christensen, F., Møller, H. (2001). The importance of head movements for binaural room synthesis, *Proc. of the 7th Meeting of the International Conference on Auditory Display*, Espoo, Finland.

Moore, Brian C. J. (2003). *An Introduction to the Psychology of Hearing*, Fifth Edition, London, UK: Academic Press.

Morrongiello, B.A., Timney, B., Humphrey, G.K., Anderson, S., & Skory, C. (1995). Spatial knowledge in blind and sighted children, *J Exp Child Psychology*, Vol. 59, pp. 211-233.

Muchnik C, Efrati M, Nemeth E, Malin M, & Hildesheimer M. (1991). Central auditory skills in blind and sighted subjects, *Scandinavian Audiology*, Vol. 20, pp. 19-23.

Nakaya,T. (1997): Statistical inferences in bidimensional regression models. *Geographical Analysis*, Vol. 29, pp. 169-186.

Nordzij, 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.

Ohuchi, M., Iwaya, Y., Suzuki, Y., & Munekata, T. (2006). A comparative study of sound localization acuity of congenital blind and sighted people, *Acoust. Sci. & Tech*, Vol. 27, pp. 290-293.

Poirier, C., Collignon, O., Scheiber, C., & De Volder, A. (2006). Auditory motion processing in early blind subjects, *Cognitive Processing*, Vol. 5, No. 4, pp. 254-256.

Röder, B., Teder-Salejarvi, W., Sterr, A., Rösler, F., Hillyard, S.A., & Neville, H.J. (1999). Improved auditory spatial tuning in blind humans, *Nature*, Vol. 400, pp. 162-166.

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*, Vol. 50, pp. 27-39.

Starlinger I. & Niemeyer, W. (1981). Do the Blind Hear Better? Investigations on Audiotory Processing in Congenital or Early Acquired Blindness, *Audiology*, Vol. 20, pp. 503-509.

Strelow, E.R. & Brabyn, J.A. (1982). Locomotion of the blind controlled by natural sound cues, *Perception*, Vol. 11, pp. 635-640.

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*, Vol. 59, pp. 1306-1328.

Vorländer, M. (2008). *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Aachen, Germany: Springer-Verlag. ISBN: 978-3-540-48829-3

Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J-P., Lepore, F. (2004). Early- and Late-Onset Blind Individuals Show Supra-Normal Auditory Abilities in Far-Space, *Current Biology*, Vol 14, pp. 1734-1738.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (1949). The precedence effect in sound localization, *Journal of Experimental Psychology*, Vol. 27, pp. 339-368.

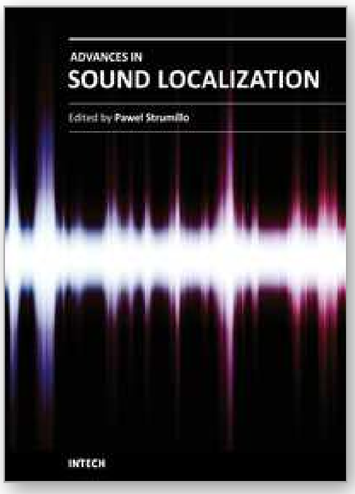
Weeks, R., Horwitz, B., Aziz-Sultan, A., Tian, B., Wessinger, C. M., Cohen, L.G., Hallett, M., & Rauschecker, J.P. (2000). A Positron Emission Tomographic Study of Auditory Localization in the Congenitally Blind, *J. Neuroscience*, Vol. 20, pp. 2664-2672.

Wenzel, E. M. (1998). The impact of system latency on dynamic performance in virtual acoustics environments, *Proc. of the 16th ICA and 135th ASA International Conference*, Seattle, WA, pp. 2405-2406.

Wightman, F. L. & Kistler, D. J. (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement, *Journal of the Acoustical Society of America*, Vol. 105, pp. 2842-2853.

Zwiers M.P., Van Opstal A.J., & Cruysberg J.R. (2001a). A spatial hearing deficit in early-blind humans, *The Journal of Neuroscience*, Vol. 21.

Zwiers, M.P., van Opstal, A.l., & Cruysberg J.R.M. (2001b). Two-dimensional sound-localization behavior of early-blind humans., *Experimental Brain Research*, Vol. 140, pp. 206-222.



Advances in Sound Localization
Edited by Dr. Pawel Strumillo

ISBN 978-953-307-224-1
Hard cover, 590 pages
Publisher InTech

Published in print edition April, 2011

Sound source localization is an important research field that has attracted researchers' efforts from many technical and biomedical sciences. Sound source localization (SSL) is defined as the determination of the direction from a receiver, but also includes the distance from it. Because of the wave nature of sound propagation, phenomena such as refraction, diffraction, diffusion, reflection, reverberation and interference occur. The wide spectrum of sound frequencies that range from infrasounds through acoustic sounds to ultrasounds, also introduces difficulties, as different spectrum components have different penetration properties through the medium. Consequently, SSL is a complex computation problem and development of robust sound localization techniques calls for different approaches, including multisensor schemes, null-steering beamforming and time-difference arrival techniques. The book offers a rich source of valuable material on advances on SSL techniques and their applications that should appeal to researches representing diverse engineering and scientific disciplines.

Brian FG Katz and Lorenzo Picinali (2011). Spatial Audio Applied to Research with the Blind, Advances in Sound Localization, Dr. Pawel Strumillo (Ed.), ISBN: 978-953-307-224-1, InTech, Available from: http://www.intechopen.com/books/advances-in-sound-localization/spatial-audio-applied-to-research-with-the-blind

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821