Forum Acusticum 2011 - Aalborg - Denmark

Acquisition of spatial knowledge of architectural spaces via active and passive aural explorations by the blind

Lorenzo Picinali
Fused Media Lab, De Montfort University, Leicester, UK.

Brian FG Katz, Amandine Afonso, and Michel Denis
LIMSI-CNRS, Orsay, France.

Summary
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, the 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 to what 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 preparing for navigation in new and unknown environments. A comparison of three learning scenarios has been performed: in-situ real displacement, passive playback of recorded navigation (binaural and Ambisonic), and active navigation in virtual auditory environment architecture. For all conditions, only acoustic cues are employed. This research is the result of collaboration between researchers in psychology and acoustics on the issue of interior spatial cognition.

PACS no. 43.55.Ka, 43.66.Pn

1. Introduction

How do blind individuals learn the configuration of a new space? What type of acoustic cues do they use to mentally represent a given environment? Is it possible for individuals to learn the configuration of a new environment on the basis of 3D Virtual Reality (VR) audio systems? These are the fundamental questions of the research project described in this paper.
The motivation of the present research was twofold. First, its aim was to contribute to documenting the processes by which blind people construct mental representations of their surrounding space. Secondly, it was intended to provide grounds for the design of systems for delivering audio information to assist blind people in their spatial orientation and navigation.

In the following sections, the different stages of the research project will be described, and the results will be analysed and discussed.
Portions of this study have been previously published [1] [2].

2. Localization versus spatial perception

Sound source localization in an anechoic environment is a unique and quite unnatural situation for most individuals. Rather, it is typical to hear sound sources with some level of reflection, even in outdoor environments, or with a high density of reflections in reverberant spaces. These additional acoustic paths from the same source can cause certain impairments, such as source localization confusion and degradation of intelligibility. At the same time, they can provide information regarding room dimensions, 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 applies more weight to the first signal that reaches the ear, i.e. the signal that comes directly from the sound source. It considers less the localization of subsequent signals resulting from reflections from walls, ceiling, floor etc. which arrive 20-40 ms after the first signal (these values can change depending on the typology of the signal, see [3]). This effect is known as the Precedence Effect [4], 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. 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 localization occurs, such as with the Franssen effect [5], but those are not the subject of this work. The later arriving signals, while not being useful for localization, are useful in interpreting the environment.

The ability to directionally analyze the early reflection components of a sound is 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, 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 exploit this information in order to gather knowledge about the spatial configuration of an environment. Elements in support of this are presented in Section 4.1, 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 localization of the direct sound, reverberation is most often simplified and 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 [6]. This study proposes to compare the late acoustic cues provided by the exploration of a real architecture with those furnished both by recordings and numerical room simulations, 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.

3. Real vs. recorded walkthrough

The path and, with it, the movements of the body and of the head of a blind person navigating in selected closed environments have been tracked and repeated, recording the experience both with a pair of binaural microphones (in-ear microphone capsules, *DPA 4060*) and with a B-Format microphone (*Soundfield ST250*).

Through binaural rendering of the recorded sound field (see Section 4.2 for more information about the conversion), and a simple player for reproducing the binaural signals, the two 3D audio recording techniques were compared through subjective evaluations. Subjects were presented the recordings and asked to describe and to construct a map of the explored space. Results were compared to those of subjects performing real explorations of the same space. Verbal descriptions performed after having listened to the recorded signals revealed a rather poor representation of the navigated environments, which was confirmed by the reconstructions, such as those shown in Fig. 2.

Due to the very poor results, indicating the difficulty of the task, the experiment was halted before all participants completed this exercise.

4. Real vs. 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 conducted with the objective of delivering information about the spatial configuration of closed environments and positions of sound sources within the environment itself, exploiting interactive virtual acoustic models.

4.1. Acoustical model

Two closed environments were selected for which 3D architectural acoustic models were created using the CATT-Acoustics (*http://www.catt.se*) software. Within each of these acoustic models, in addition to the architectural elements, different sound sources from the real situation (both natural and installed audio loops) were included in order

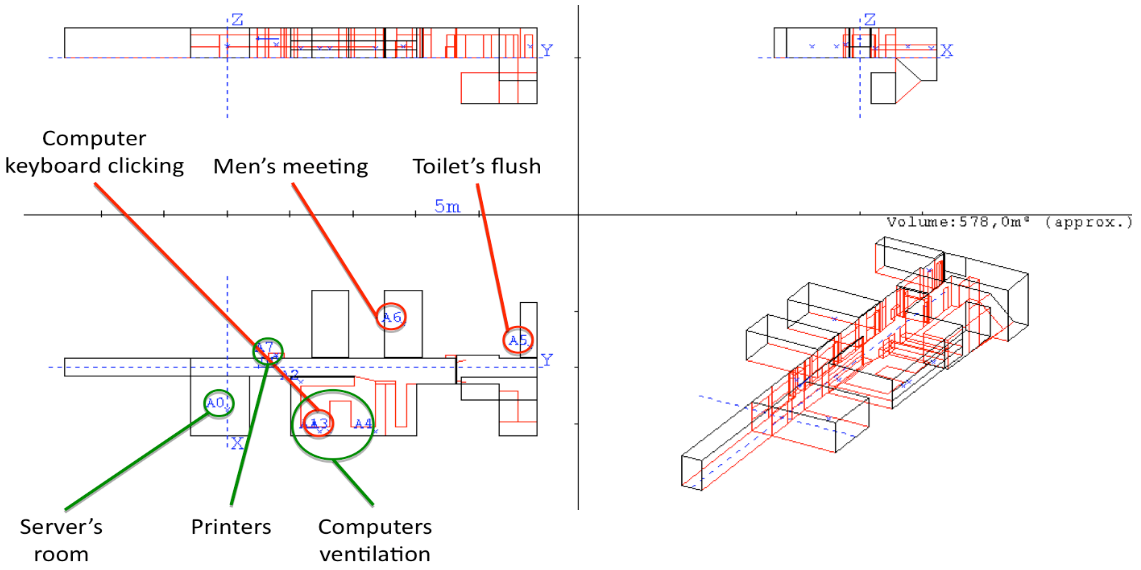

Figure 1. Geometrical acoustic model of one of the experimental spaces, including positions of real (**green** lines and circles) and installed audio loop playback (**red** lines and circles) sources.

to be able to carry out a distance comparison task (see Section 4.5). 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. 1.

After observing in the real navigation stage 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.) by listening to the reflections of the acoustic signals (see also Section 2), 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 2nd order Ambisonic rendering engine was used (as opposed to the pre-recorded walkthrough which was 1st order) to improve spatial precision while still allowing for dynamic head rotation.

4.2. The VR navigation platform

Due to the large number of concurrent sources and the size of 2nd order impulse responses (IR), a real-time accurate rendering was not feasible. Therefore, an alternate approach was elaborated in order to achieve real-time results. As a first step, navigation was limited to one dimension only. Due to the fact that the test 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 to be able to simulate footfall noise and waist height to simulate finger snapping noise.

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, and panning between the different signals during the real-time navigation was employed, rather than performing all convolutions in real-time. Finger snap and footfall noises were rendered in real-time and added to the soundscape.

The final 2nd order Ambisonic signal was then rendered over binaural headphones employing the approach of virtual speakers. The 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 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, converting a monophonic signal to a stereophonic binaural signal. 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) from the Listen IRCAM database (*http://recherche.ircam.fr/equipes/salles/listen/*). More information about this approach can be

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

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

found in [7]. Full-phase HRIRs 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 participant, and an HRTF selection phase, were also performed so that an optimal binaural rendering could be performed.

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. 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.

4.3. Protocol: real vs. 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 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. Three congenitally blind and two late blind individuals (three females, two males) participated in the virtual condition.

Assessment of the spatial knowledge acquired in the two learning conditions involved two evaluations: a reconstruction of the environment using Lego® blocks and a test concerning the mental comparison of distances. The order of the tests was reversed between the two corridors.

4.4. Lego reconstructions analysis

Several measures were made on the resulting block reconstructions: number of sound sources indicated, 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. 2). Door openings into rooms containing a sound source were well identified, while more difficulties were found with those with no sound source present. Participants were also capable of distinctively perceive the various surface material changes along the corridors. An objective evaluation of how similar the different reconstructions are from the actual map of the navigated environment was carried out using bidimensional regression analysis [8]. After some normalisation, the positions of the numerous reference points, both architectural elements and sound sources comprising a total of 93 coordinates, were compared with the corresponding point in the reconstructions, with a mean number of 46±12 points over all subjects. The bidimensional regression analysis resulted in a correlation index between the true map and the reconstructed map. Table 1 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 improvement in the quality of the mental reconstruction is confirmed by the similar values in terms of map correlation. 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.

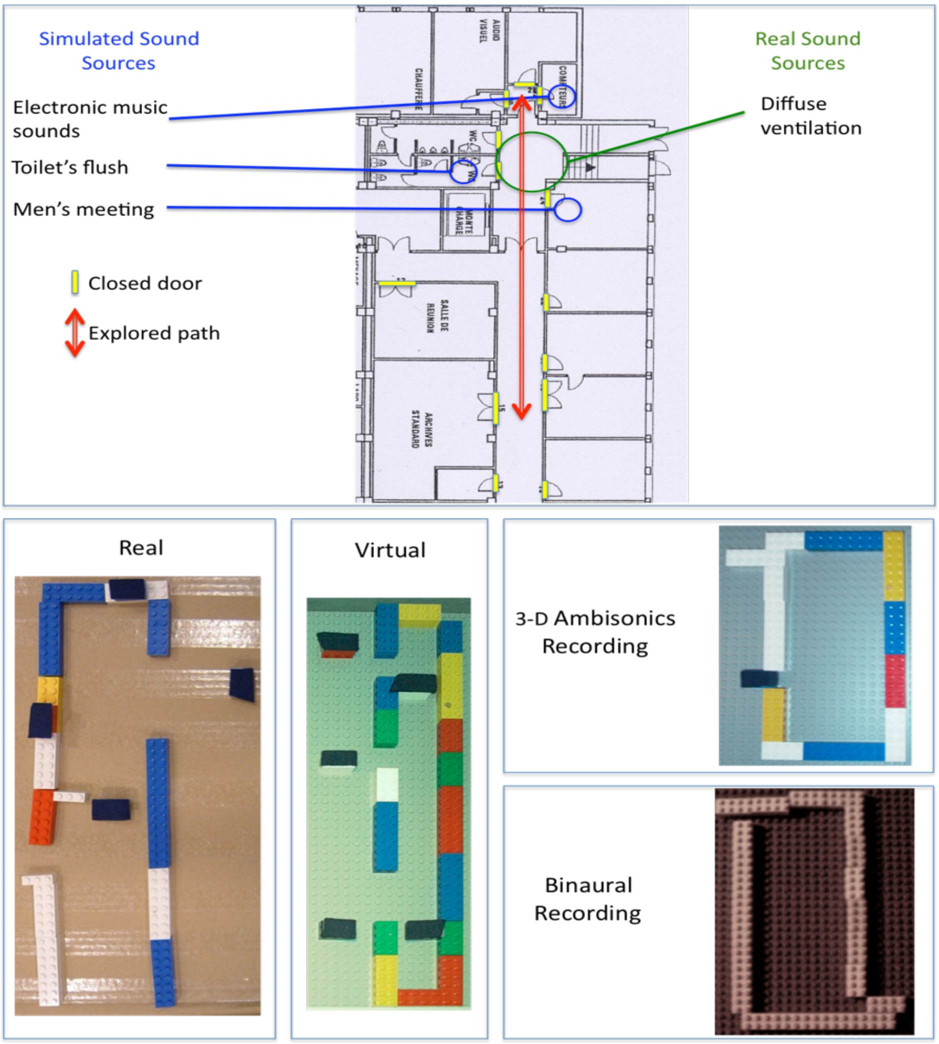

Figure 2. Examples of LEGO reconstructions following real navigation, virtual navigation, binaural and Ambisonics playback

4.5. Distance comparison analysis

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 response latency is longer for smallest differences. The *symbolic distance effect* is taken as reflecting the analog character of the mental representations and the capacity of preserving actual distance metrics [9].
In addition to the starting and arrival points, three sound sources existed along each path within the two navigated environments (1^{st}: keyboard, men's voices and toilet flush; 2^{nd}: 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, large. Participants were presented with two distance pair verbally, and asked to indicate which of the two was longer.
Analysis of the results focused on the frequency of correct responses. Table 2 shows the frequency of correct responses for the participants for both real and virtual and real navigation conditions.
Results show that even with a high level of performance for the real navigation condition,

Table II. Percent of correct responses for distance comparisons as a function of navigation condition

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

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 nearly 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 that preserved topological and metric properties of the original environment.

5. Conclusions and future work

Some interesting comments 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 2, 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 solely through auditory exploration.

These results are of special value by showing that the behavioral measures did not differ whether people navigated in actual or virtual corridors. Navigating with the help of a joystick makes it possible to build a mental representation of the spatial structure of an environment populated with sounds. This mental representation preserves the topological and metric properties of the environment, just as it does after actual navigation.

Possible future development for similar applications could be linked to the design of a 3D VR acoustical model specifically customized for the blind. In the present research, virtual acoustic software normally used for sighted people was employed, whilst it is knows that the auditory perception of blind people can be largely different, for example in terms of environmental comprehension and localization accuracy.

Extensive behavioural tests should be carried out, on simpler environmental configurations, in order to perform a “fine tuning”, specifically for applications for the blind, of all the acoustical characteristics of the simulated space.

Acknowledgements

This study was supported by a grant from the European Union (STREP Wayfinding, Contract 12959). 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*).

References

[1] M. Denis, A. Afonso, L. Picinali, B. FG Katz: Blind people’s spatial representations: Learning indoor environments from virtual navigational experience. Proc. of the 11th European Congress of Psychology, 7-10 July 2009, Oslo, Norway.

[2] B. FG Katz, L. Picinali: Spatial Audio Applied to Research with the Blind. Strumillo, P., ed. Advances in Sound Localization, INTECH, 2011, *(in press)* ISBN: 978-953-307-224-1.

[3] Brian C. J. Moore: An Introduction to the Psychology of Hearing, Fifth Edition, pp 253-256. Academic Press, London, UK, 2003.

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

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

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

[7] A. McKeag, D. S. McGrath: Sound Field Format to Binaural Decoder with Head Tracking. Proc. of the 101st AES Convention, 1996, Los Angeles, CA.

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

[9] M. Denis: 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 (2008), 197-210.