EXPLORATION OF VIRTUAL ACOUSTIC ROOM SIMULATIONS BY THE VISUALLY IMPAIRED

Reference PACS: 43.55.Ka, 43.66.Qp, 43.55.Hy

***Katz, Brian F.G.*[1] *;Picinali, Lorenzo*[2]**

[1]*LIMSI-CNRS, Orsay, France.*
brian.katz@limsi.fr
[2]*Fused Media Lab, De Montfort Univ., Leicester, UK.*
LPicinali@dmu.ac.uk

ABSTRACT

Virtual acoustic simulations of two interior architectural environments were presented to visually impaired individuals. Interpretations of the presented acoustic information, through block map reconstructions, were compared to reconstructions following in-situ exploration as well as playback of binaural and Ambisonic walkthrough recordings of the same spaces. Results show that dynamic exploration of virtual acoustic room simulations outperforms passive recording playback situations, despite dynamic rotation cues offered by Ambisonic playback. Simulations used off-line HOA RIR synthesis and a hybrid rendering combining pre-convolved signals and real-time convolutions for sounds related to user displacement and self-generated noise.

1. CONTEXT

The use of virtual acoustics auralization has become common practice in recent years in the domain of architectural acoustic consulting. Such simulations are also often used in historical and acoustical archaeological studies, offering audio results for aesthetic judgments or qualitative comparisons on standardized parameters. However, these types of applications rarely focus on the true realism of the simulation from a non-aesthetic point of view. In contrast, this study presents a study where virtual room acoustic simulations are presented to visually impaired individuals to evaluate if these simulated environments are sufficiently accurate so that visually impaired users can correctly describe the architectural space. If successful, such a virtual simulation could be used as an aid for visually impaired individuals to learn the configuration of new and unknown spaces. For example, upon taking a new job in a new building, the individual could explore the building at home so as to be able to move more freely once on-site. In addition, details of room acoustic simulation calculation and rendering methods can be explored relative to which acoustic cues are more pertinent for spatial acoustic perception. With such information, virtual acoustic simulations for the visually impaired could be improved, by refining such cues, and optimized, by reducing calculation costs for non-relevant cues. In the following sections, the different stages of the study will be described. Special attention will be paid to the system rendering architecture for real-time navigation.

2. OVERVIEW

Various studies have attested to the capacity of the blind to navigate in complex environments without relying on visual inputs [3][7]. In the absence of sight, kinesthetic experience is a valid alternative source of information for constructing mental representations of an environment. Typical protocols consist of participants learning a new environment by locomotion (with or

without a guide), followed by various mental operations on their internal representations of the environment. For instance, participants could estimate distances and directions from one location to another one [3].

Recent studies have employed virtual auditory reality simulations to investigate the role of the learning experience in the acquisition of spatial knowledge by blind people (see [2]). Active exploration in the virtual environment was compared to verbal descriptions. When participants performed localization tasks (pointing towards the location of different targets within the environment), errors were higher with the verbal description group. Furthermore, following a mental distances comparison task between pairs of targets, response times confirmed that longer distances systematically required longer scanning times, reflecting that the metrics of the original scene were preserved in the internal representation of the environment [1].

Most interactive systems (e.g. gaming applications) are visually-oriented. While some engines take into account source localization of the direct sound, reverberation is 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 [10]). The study presented in the following sections proposes to compare the acoustic cues provided by a real architecture with those furnished both by *in-situ* 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 systems specifically created and calibrated for visually impaired individuals.

In contrast to previous studies, this work focuses primarily on the understanding of an architectural space, and not on the precise localization of sound sources. As a typical case, this study was performed in two corridor spaces in a laboratory building (see Fig. 1). These spaces are not exceptionally complicated, containing an assortment of doors, side branches, ceiling material variations, stairwells, and static noise sources. In order to provide reference points for certain validations, some additional sound sources were added using simple audio loops played back over portable loudspeakers. Results for distance comparison tasks have been previously presented elsewhere (see [4][5][9]). This paper presents the technical aspects of the *in-situ* recordings and the virtual room acoustics simulations. Results concerning map reconstructions of the spaces are also presented.

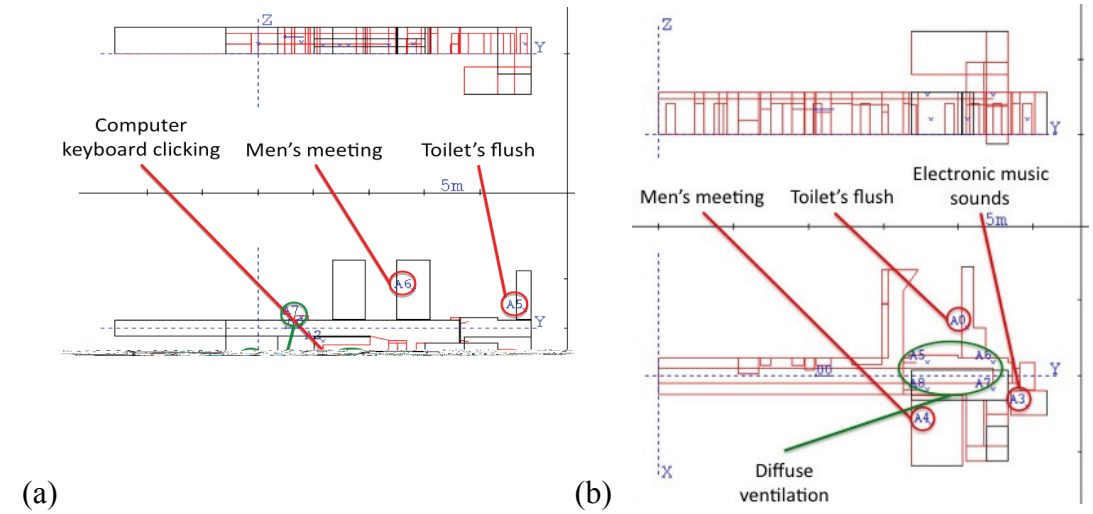

*Figure 1. Geometrical acoustic model of the (a) first and (b) second experimental spaces, including positions of real (**green** lines and circles) and installed audio loop playback (**red** lines and circles) sources.*

3. IN-SITU 3D RECORDING

3.1 Recording method

Recordings were carried out in two different sessions. In the first session, a blind person equipped only with in-ear binaural microphones navigated the environment while his path, body, and head movements were tracked via multiple synchronized CCTV cameras and a system of

markers positioned along the walls of the environment. No white-cane or guide-dog was allowed, and the individual was asked to avoid contact with any surfaces, and recommended to remain along the centerline. No walking speed or head movements were imposed, and he was asked to make any movements or noises as necessary to obtain a confident sense of the architectural space. The subject made one down and back trip for each corridor; no contact was ever made with any wall or other object.

Subsequent to this, in a second session, an operator equipped with both binaural and B-format microphones precisely repeated the trajectories. The path, movements, and any self-generated noises (other than commentaries) were reconstructed following a precise timeline established from analysis of the first session's recording.

3.2 Playback method

Two methods were employed in order to reproduce the different recorded signals. For the binaural playback, a simple stereo player was used. In the case of the B-Format recording, a conversion to binaural was necessary. The 1st order recorded Ambisonic signal was rendered over binaural headphones using the virtual speaker approach. 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 use of head tracking allowed for the orientation of the 3D sound-field to be modified in real-time, performing rotations in the Ambisonic domain as a function of participant's head movements, thereby keeping the scene stable in the world reference frame. The rotated signal was then decoded on a virtual loudspeakers system with the sources placed on the vertices of a dodecahedron. These twelve decoded signals were then rendered as individual binaural sources via twelve instances of a binaural spatialization algorithm, converting each monophonic signal to a stereophonic binaural signal. The twelve binauralized virtual loudspeaker signals were then summed and presented to the subject.

The binaural spatialization algorithm used [6] employs time domain convolution with Head Related Impulse Responses (HRIR) from IRCAM's Listen project database (http://recherche.ircam.fr/equipes/salles/listen/). More information about this approach can be found in [5]. Full-phase HRIRs were used, rather than minimum-phase simplifications, in order to maintain a highest level of spatial information. Customization of the Interaural Time Differences (ITD), using a head circumference model of the participant, and an HRTF selection phase, were also performed to improve so that an optimal binaural rendering could be achieved.

3.3 Exploration protocol

Each subject was presented with one of the two types of recordings for each of the two environments. Participants were seated during playback. The initial part of the session comprised a learning phase, consisting of repeated listenings to the playback until the participants felt they understood the environment. In the binaural recording condition, participants were totally passive, and instructed to remain still with a fixed head orientation. Acoustic cues related to head movements and orientation in the scene were dictated by the state of the operator's head during the recording. In the Ambisonic recording condition, participants were able to feely 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.

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

4. VIRTUAL ROOM ACOUSTIC SYNTHESIS

A key element observed in the in-situ recording playback condition was the lack of interactivity and free movement within the simulated environments. Discussions with the initial participants of the in-situ experimental condition highlighted this fact, and the difficulty in interpreting the recordings. Through the use of an interactive virtual environment, it was hoped that this issue could be addressed, at least to some degree. While interactivity is more feasible in a virtual simulation, the accuracy of the numerical simulations and the complexity or richness of the audible soundscape may be more limited. For this initial study, a truly interactive real-time room acoustic simulation was considered too costly in computational resources. As such, a hybrid simulation was developed, combining off-line calculated room impulse responses (RIR) and convolutions with real-time panning and mixing.

4.1 Acoustical model

3D architectural acoustic models were created for the two corridors using the CATT-Acoustics software (http://www.catt.se). Within each of these acoustic models, in addition to the architectural elements, the different sound sources from the real situation were included in order to present a comparable scene. A third, simple geometry model was also created for a training phase, so that subjects could become familiar with the overall interface and exploration protocol. The geometrical models of the two experimental spaces are shown in Fig. 1. Acoustical surface material definitions were determined and adjusted iteratively so as to match the materials present in the real environments. RIR measurements were performed in two positions for each environment (one position in the middle and one at the far end, near the staircase). The simulation's material definitions were adjusted so as to obtain the same room acoustical parameters, RT60 in octave bands, between the simulated and measured RIR.

It was observed in the real navigation phase that blind individuals made extensive use of self-generated 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. As such, the 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 (HOA) rendering engine was used (as opposed to the pre-recorded walkthrough using 1st order) to improve spatial precision while still allowing for dynamic head rotation.

4.2 The navigation platform

Due to the large number of concurrent sources and to the size of HOA RIRs, a real-time accurate rendering was not feasible. A more economical yet high performance hybrid method was developed. As a first step, navigation was limited to one dimension only. Benefiting from the fact that both environments were corridors, the user was restricted to movements along the centerline. Receiver positions were defined at equally spaced positions (every 50 cm) along this line, at head height. The different noise source positions, as indicated in Fig. 1, were included, providing a collection of HOA RIR for each receiver position. In order to provide real-time navigation of such complicated simulated environments, a pre-rendering of the HOA signals for each position of the listener was performed off-line using in-situ recordings or the same audio file loops as were used in the real condition. At navigation time, a simple Ambisonic panning was performed between the nearest points along the centerline pathway, rather than performing all convolutions in real-time.

To include self- generated noises, source positions at ground level (for footfall noise) and waist height (finger snap noise) were also included. Finger snap and footfall noises were rendered off-line and added in real-time to the final Ambisonic soundscape. The final Ambisonic mix was converted to binaural using the same approach described in Section 3.2, though extended to account for 2nd order Ambisonic format.

In the experimental condition, participants were provided with a joystick as a navigation control device and a pair of headphones equipped with the head-tracking device (as in Section 3.2). Footfall noise was automatically rendered in accordance with the participant's displacement in

the virtual environment, approximating a 50 cm stride. The navigation speed was continuously variable from 0.1 to 1 m/s, proportional to the degree of forward pressure applied to the joystick. The finger snap was played each time the listener pressed a button on the joystick.

In total, 44 receiver positions were calculated for the first corridor and 31 for the second. As can be seen in Fig. 1, for the first environment 4 virtual sound sources were created for simulating the real sources in the real environment (2 sources were used for simulating the computer ventilation noise), while an additional 3 virtual sources were created for simulating the artificial looped audio playback sources. Similarly for the second corridor, 4 real sources (used to simulate the diffuse ventilation noise) and 3 artificial ones were defined. In both virtual spaces, a total of seven HOA source-receiver pair RIRs were synthesized for each receiver position (308 and 217 RIRs for the first and second corridor respectively). In addition, for each receiver position, a corresponding RIR was synthesized for simulating the finger snapping noise: the source, in this case, was different for each receiver, positioned at a height of 110 cm and a distance of 50 cm ahead of the receiver in order to more accurately represent the position of the hand's location. Finally, to account for the footstep noise, a RIR was synthesized for each receiver position at a height of 1 cm, and at a distance of 10 cm to the left of the centerline for the left step, and correspondingly to the right for the right step. The step side was alternated, starting with the right foot forward. A total of 396 HOA RIRs were synthesized for the first corridor and 279 for the second.

Each RIR was pre-convolved with the corresponding audio source signal. For the real sources, signals have been recorded in the real environment (as close as possible to the noise sources in order to minimize acoustical contributions of the room in the recordings). For the virtual sources, the same signals used for the audio playback loops in the real navigation condition were used. Two audio samples were selected for the finger snap and footstep noise, allowing for source variation. A total of 3564 and 2511 signals were convolved for the first and second corridor respectively.

The convolved HOA signals corresponding to the seven static sources were summed for each receiver's position. The resulting 9-channel mixes were then read in a MaxMSP patch and played back synchronously in a loop (the length of the signals was approximately 2 minutes). In order to make the processing more efficient, the multichannel player only played the signals corresponding to the two receiver positions closest to the actual position of the individual during the virtual navigation. A cosine-based crossfade was performed between these two HOA signals relative to the position. The playback of the convolved signals of the finger snapping noise was activated when the individual pressed one of the buttons on the joystick, cross-faded in a similar fashion. The footstep noise, with the position chosen relative to the current navigation position, was played at every displacement interval of 50 cm without any cross-fade. The resulting HOA 9-channel audio stream, comprising the sum of the static sources, finger snapping, and footstep noise, was then sent to the virtual loudspeakers conversion algorithm as previously described.

5. EVALUATION

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. The assessment of spatial knowledge acquired in the two conditions was examined through the creation of a map reconstruction of each environment. A distance comparison task was also performed (for results, see [1]). 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.

Two congenitally blind and three late blind participants (two female, three male) took part in the in-situ recording condition. Verbal descriptions for the in-situ recording condition revealed that participants acquired a rather poor understanding of the navigated environments. This was further confirmed by analysis of the reconstructions. Fig. 2 shows reconstructions of the second corridor space for the different 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 confirmed by the verbal description. The reconstruction shown following the Ambisonic playback condition reflects similar poor and misleading mental representations. Due to the very poor results for this test, indicating the difficulty of the task, this experiment was terminated prior to any additional participants completing the experiment.

In the virtual condition, three congenitally blind and two late blind individuals (three females, two males) explored the same two corridors. As a reference condition, two congenitally blind and three late blind individuals (three females, two males) performed the exploration and reconstruction task via real exploration for the two corridors.

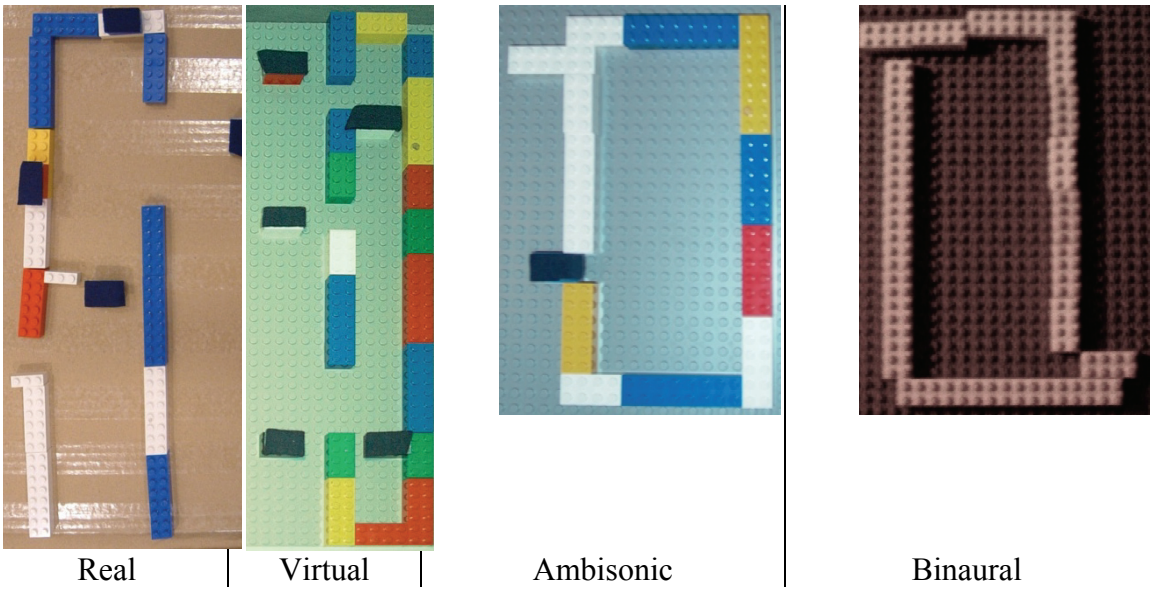

Figure 2: Photographs of representative map reconstructions of the second corridor space following real navigation, virtual navigation, Ambisonic playback, and binaural playback.

5.1 Maps

The map reconstructions made by each participant were photographed. Each map also included a corresponding audio description by the participant, to help understand the elements used. An example of a reconstruction for each exploration condition is shown in Fig. 2. 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 were 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 greater difficulty was observed for rooms with no sound source present. Participants were also capable of distinctively perceiving the various surface material changes along the corridors.

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 comparably poorer localization accuracy provided by this particular recording/restitution technique. Interestingly, participants in the playback conditions failed to comprehend that the recordings were made in a straight corridor with openings on the two sides.

In order to perform more detailed comparisons, each map was manually transcribed into MatLab in order to create a numerical version. A reference template (see Fig. 3(a) for the first corridor template) was created which included all sound sources and the basic architectural dimensions. A total of 93 different coordinate elements were included for the first corridor, while

only 46 were include for the simpler second corridor. Examples of numerical map reconstructions for different conditions are shown in Fig. 3. From analysis of the reconstructions, participants identified a mean of 46±12 points for the first corridor and 20±9 for the second.

5.2 Correlation

An objective evaluation, as opposed to a human visual comparison, on how similar the different reconstructions are from the actual maps of the navigated environments was carried out. A 2D bidimensional regression analysis [8] provided a correlation index between the reference map and each reconstructed map. This method included some normalization in order to account for different scales used between participants, as well as for the reference map. Only those elements present in each individual's reconstruction were used in the correlation computation. This resulted in a bias for maps with very few identified elements, as for example a simple rectangle would have a high correlation index, but would not represent a high degree of understanding for the space. Results for the correlation analysis of all conditions and subjects are shown in Fig. 4 with respect to the relative number of identified elements in each map reconstruction.

While the number of subjects in the experiment is relatively low, due to time and conditions necessary for both the environment and the participant, there are some clearly observable tendencies. In the real exploration condition, both the correlation index and the quantity of identified elements are rather high. For both corridors, the Ambisonic and Binaural conditions present rather low correlations and number of identified elements, relative to the other conditions. As no participant performed more than one condition, there are likely to be some individual variance effects, but all subjects expressed their comfort in the task and their understanding of the space at the time of the experiment.

There was a notable difference in the results between the two corridors, with the virtual condition in general providing higher correlation values than even the real condition. In contrast, in the second corridor, while the virtual condition still performed generally better than the two *in-situ* conditions, it was not comparable to the real condition. Due to the limited number of participants, no analysis was performed comparing the results between early and late blind individuals.

6. CONCLUSION

Overall, results showed that listening to passive binaural playback or Ambisonic playback, which also included interactive head-movements, provided less usable information than a virtual simulation with respect to the acquisition of spatial information of an interior architectural environment. The presence of both dynamic cues relative to displacement and controlled events such as finger snaps, as included in the virtual condition, were deemed highly valuable by the participants. Virtual acoustic simulations provided acoustic information that allowed for highly correlated detailed map reconstructions relative to a real exploration condition. Some differences were found between the two experimental corridors, with the more complex environment offering better results than the corridors with more diffuse noise sources.

7. ACKNOWLEDGEMENTS

This study was supported in part by a grant from the European Union (STREP Wayfinding, n° 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*).

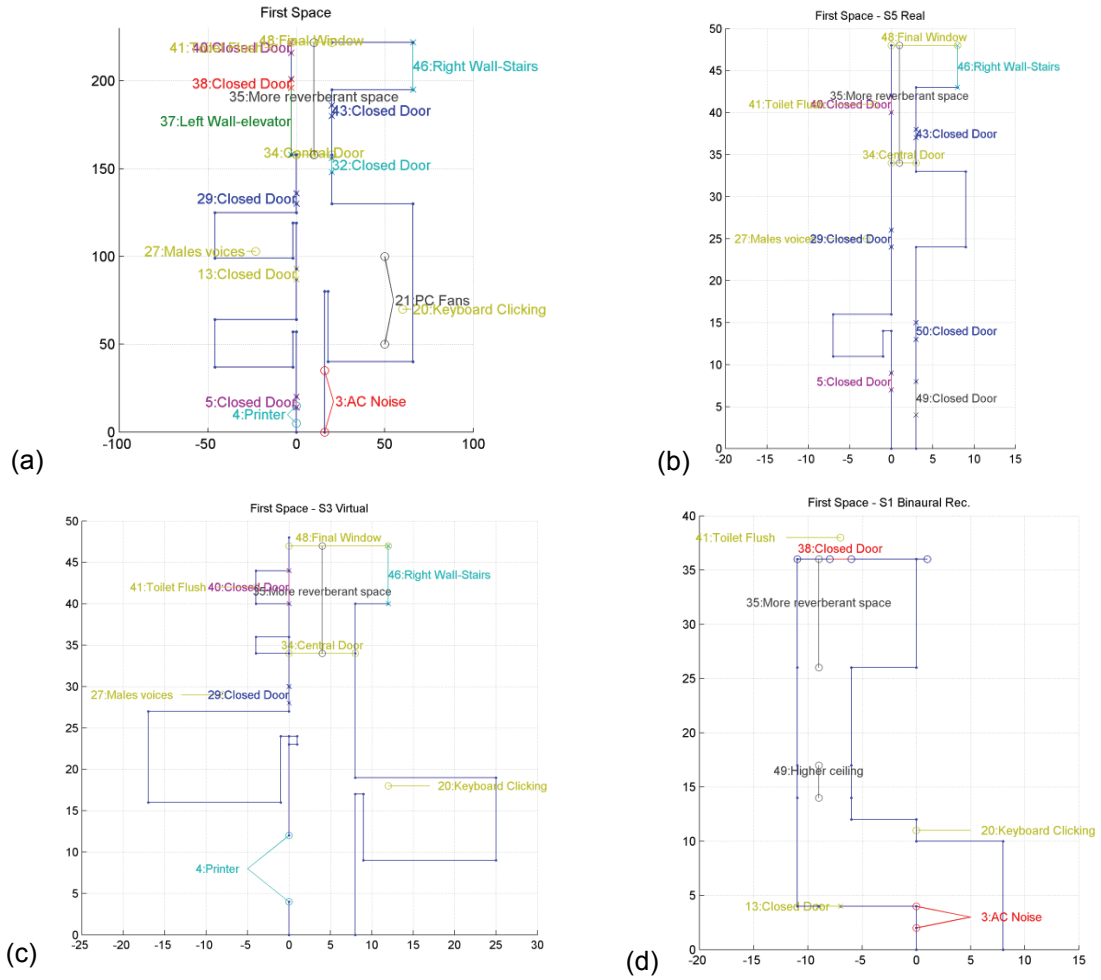

Figure 3. (a) Reference map for the first corridor space, units in decimeters. Example transcribed map reconstructions for (b) real, (c) virtual, and (d) binaural recording exploration conditions, units in LEGO pips.

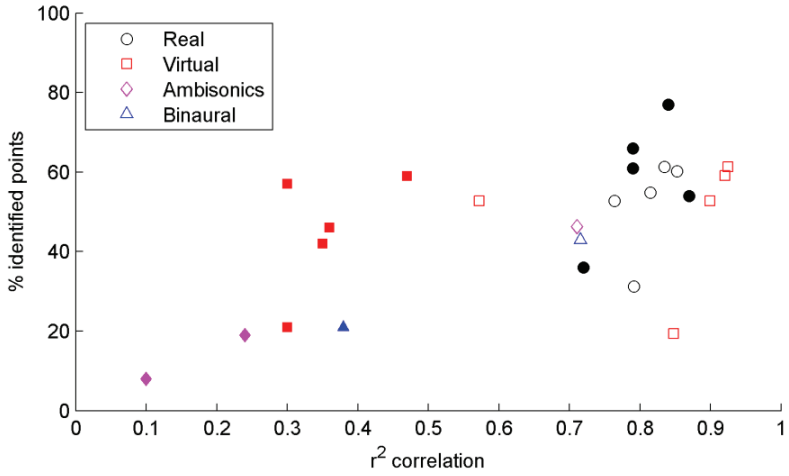

Figure 4. Correlation index versus percentage of identified elements in the map reconstruction for all subjects for the first (open markers) and second (filled markers) corridor spaces.

8. REFERENCES

[1] 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.

[2] Afonso, A., Katz, B. F. G., Blum, A. & Denis, M. (2005). *Spatial knowledge without vision in an auditory VR environment*, Proc. of the XIV meeting of the European Society for Cognitive Psychology, Leiden, the Netherlands.

[3] Byrne, R. W. and Salter, E. (1983). Distances and directions in the cognitive maps of the blind, *Canadian Journal of Psychology*, 70.

[4] Denis, M., Afonso, A., Picinali, L. & Katz, B.F.G (2009). *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.

[5] Katz, B.F.G. & Picinali, L. (2011) Spatial Audio Applied to Research with the Blind. *Advances in Sound Localization*, Strumillo, P., ed., INTECH, 2011.

[6] LSE (2010) IDDN.FR.001.340014.000.S.P.2010.000.31235 LSE (LIMSI Spatialization Engine)

[7] Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., & Fry, P. A. (1993). Nonvisual navigation by blind and sighted: Assessment of path integration ability. *Journal of Experimental Psychology*: General, 122.

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

[9] Picinali, L., Katz, B.F.G., Afonso, A. & Denis, M. (2011). *Acquisition of spatial knowledge of architectural spaces via active and passive aural explorations by the blind*. Proc. of the Forum Acusticum 2011, Aalborg, 27-June - 1-July, 2011.

[10] Vorländer, M. (2008). *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer-Verlag, Aachen, Germany, 2008.