Int. J. Human-Computer Studies 85 (2016) 4–15

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Interactive spatial sonification for non-visual exploration of virtual maps ☆

Michele Geronazzo [a,*], Alberto Bedin [a], Luca Brayda [b], Claudio Campus [b], Federico Avanzini [a]

[a] Department of Information Engineering, University of Padova, Via Gradenigo 6/B, 35131 Padova, Italy
[b] Department of Robotics, Brain and Cognitive Sciences, Fondazione Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy

ARTICLE INFO

Article history:
Received 21 January 2015
Received in revised form
9 July 2015
Accepted 13 August 2015
Available online 1 September 2015

Keywords:
Multimodal interaction
3D audio
Binaural sound
Virtual maps
Non-visual navigation
Haptics
Visual impairment
Multisensory integration
Haptic mouse

ABSTRACT

This paper presents a multimodal interactive system for non-visual (auditory-haptic) exploration of virtual maps. The system is able to display haptically the height profile of a map, through a tactile mouse. Moreover, spatial auditory information is provided in the form of virtual anchor sounds located in specific points of the map, and delivered through headphones using customized Head-Related Transfer Functions (HRTFs). The validity of the proposed approach is investigated through two experiments on non-visual exploration of virtual maps. The first experiment has a preliminary nature and is aimed at assessing the effectiveness and the complementarity of auditory and haptic information in a goal reaching task. The second experiment investigates the potential of the system in providing subjects with spatial knowledge: specifically in helping with the construction of a cognitive map depicting simple geometrical objects. Results from both experiments show that the proposed concept, design, and implementation allow to effectively exploit the complementary natures of the "proximal" haptic modality and the "distal" auditory modality. Implications for orientation & mobility (O&M) protocols for visually impaired subjects are discussed.

1. Introduction

Our everyday interactions with the world are intrinsically multimodal. Assessing the relevance of a given sensory modality on user behavior and performance is not trivial. In fact, the relevance of a given modality depends on the accuracy and exploitability of the information that it carries (Gibson and Pick, 2000). Moreover, multiple modalities combine through integration and combination (Ernst and Bulthoff, 2004). They interact with each other also by means of cross-modal effects and correspondences (Driver and Spence, 2000; Spence, 2011). A thorough understanding of these mechanisms is an essential asset for the development of human–computer interfaces in which several unimodal streams of information merge effectively into a multimodal interactive display.

Spatial perception and spatial cognition (i.e., the acquisition, organization, utilization, and revision of knowledge about spatial environments) rely on multimodal information, although different modalities differently contribute to the perception and cognition of space. More precisely, sensory modalities are traditionally classified into "distal" and "proximal". Touch is typically considered a proximal sense because tactile stimuli are generated as a consequence of a direct contact between the body and external objects. In contrast, vision is a distal sense as stimuli in vision arise from distant objects. In the absence of vision, audition is the next most reliable distal sense available. Accordingly, a largely accepted view (originally attributable to Révész (1950)) is that the perception of space differs between modalities: haptic space is centered on the body, whereas vision and audition are centered on external coordinates.

A partially related concept (see Loomis et al., 2001 for an extended discussion) is that spatial cognition, and especially navigation within an environment, is based on two distinct and complementary mechanisms: landmark-based navigation and path integration. Landmarks provide subjects with direct sensory information about their current position and orientation, possibly in conjunction with an external or cognitive map, while path integration allows them to constantly update their perceived position and orientation, e.g. relative to some starting point.

Sensory modalities differ not only in terms of sensory input channels or neural activity involved, but also in terms of sensorimotor skills. The differences between "hearing" and "seeing" lie in the fact that, among other things, one is seeing if there is a large change in sensory input when blinking; on the other hand, one is

☆This paper has been recommended for acceptance by E. Motta

* Corresponding author. Tel.: +39 049 827 6976.
E-mail addresses: geronazzo@dei.unipd.it (M. Geronazzo), bedinalb@dei.unipd.it (A. Bedin), luca.brayda@iit.it (L. Brayda), claudio.campus@iit.it (C. Campus), avanzini@dei.unipd.it (F. Avanzini).

http://dx.doi.org/10.1016/j.ijhcs.2015.08.004
1071-5819/© 2015 Elsevier Ltd. All rights reserved.

hearing if nothing happens when one blinks but there is a left/right difference when one turns the head, and so on. This line of reasoning leads to the concept of it sensory substitution and specifically implies that it is possible to obtain a visual experience from auditory or tactile input, provided the sensorimotor laws that are being obeyed are the laws of vision O'Regan and Noë, 2001. Starting in 1967, the pioneering work of Bach-y-Rita on tactile-to-vision substitution (see Bach-y Rita, 2004 for a review) first showed the potential of this approach.

Touch and audition have been used in several applications aimed at improving orientation and mobility (O&M) performance, especially for visually impaired users, and particularly at supporting spatial knowledge at least at three different levels: knowledge about a point in space (e.g., a landmark or a destination), knowledge about a sequence of points (i.e. a path to a destination, or "route knowledge"), and integrated knowledge about the environment (i.e., cognitive-map like knowledge, or "survey knowledge"): see Wiener et al. (2009) for a detailed discussion. In everyday life subjects gain "on-line" cognitive mapping through visual information while exploring an unknown environment, and additional "off-line" information (maps and pictures) is often previously available. Visually impaired users, who lack the possibility to collect these two types of information, are required to use compensatory sensory channels. In particular, obtaining substantial information and mapping of an unknown space before arriving to it has been recognized to be crucial for supporting secure O&M performance (Lahav and Mioduser, 2008; Loeliger and Stockman, 2014).

One further issue is the possible dependence of internal spatial representations and processes on previous and current visual abilities (e.g., late-onset vs. early-onset blindness and total blindness vs. low-vision). Although blind subjects perform less efficiently of sighted subjects in many tasks requiring imagery, this is not always the case and there is evidence that they often employ compensatory mechanisms can overcome the limitations of sight loss (Cattaneo et al., 2008). Moreover, some researchers suggest that spatial representations are partially amodal, i.e. independent from encoding modalities (Loomis et al., 2002; Klatzky et al., 2003). This perspective is supported by experimental results showing similarities between learning modalities and between groups of sighted, late-blind, and early-blind users (Afonso et al., 2010).

To date the most used navigation aid in the domain of O&M activities is still the white cane, which (i) requires the user to be immersed in the exploration space, (ii) provides mainly proximal haptic information, collected through continuous interaction with the environment, and (iii) provides proximal and distal information implicitly emerging in the propagation of continuous impact sounds. However, recent literature has shown that the development of *ad hoc* multimodal virtual environments that allow active exploration by users can be a great help for the acquisition of spatial knowledge. Besides the already mentioned pioneering work by Bach-y-Rita and coworkers, several applications appeared in the last two decades, thanks to the increased computational and expressive power. 3D audio in particular has been increasingly exploited, as demonstrated by the introduction of the word "auralization" back in the early 1990s (Vorländer, 2008). Also, a recent systematic review of the literature on the sonification of physical quantities (Dubus and Bresin, 2013) shows that spatial features of sound are particularly effective to sonify quantities related to kinematics, i.e. relative to motion and position.

Already Usoh et al. (1999) presented a comprehensive study of the application of 3D audio and non-speech based sounds for navigation and path finding. Another system for virtual map exploration, *HOMERE* (*Haptic and audiO Multimodality to Explore and Recognize the Environment*) (Lecuyer et al., 2003), replicated a virtual white cane used during exploration of predefined paths, and required a large and expensive installation with an audio surround system. Lahav and Mioduser (2004, 2008) performed several studies on map exploration by blind subjects, providing insights about exploration strategies, methods, and processes, as well as guidelines about the main features required for the development of haptic virtual environments: in particular their results showed a reduction in the exploration time for participants previously trained in a virtual space compared to control groups who explored a real space without training. These results led to the development of a complex desktop virtual reality environment, *BlindAid* (Schloerb et al., 2010; Lahav et al., 2012), based on a Phantom®[1] device and spatial audio, and employing exploration actions and commands borrowed from standard human–computer interaction commands (such as zooming, scrolling, pausing, and undoing).

A comprehensive set of works on the use of 3D audio for O&M applications is due to Walker and coworkers. In a study on the effects of non-speech auditory beacons in navigation performances, they used sounds that change in timbre and position with increasing levels of subject's practice, and showed that the learning experience of the system improved both speed and accuracy in navigation performances (Walker and Lindsay, 2006). More recent studies focused specifically on systems for the understanding of the layout of an unfamiliar room (Jeon et al., 2012) and the memorization of spatial scenes by means of 3D audio and idiothetic cues (Viaud-Delmon and Warusfel, 2014). Katz and coworkers also provided a substantial amount of research to this topic. In particular, they developed a system exploiting a 3D audio virtual environment to investigate structural properties of spatial representations in visually impaired people, initially presented in Afonso et al. (2010) and Katz et al. (2012a). They also used a "ears in hand" metaphor, originally proposed by Magnusson et al. (2006): this metaphor proposes an egocentric view of the virtual map, in which the ears of the user are virtually placed at the position of the hand or handheld haptic device used to explore the map. Their experimental results showed that the use of this metaphor was found to be intuitive without any prior training (Menelas et al., 2014).

Taken together, these works show that a proper combination of haptic and auditory modalities can convey relevant information for off-line learning of virtual setups that mimic real environments. More precisely, the multimodal feedback should be designed so that intermodal conflicts are minimized, that information is consistent with a real and ecological representation of the world, and that the cognitive load for the user remains low. Auditory cues in particular can provide useful global spatial information while haptic feedback can be exploited to guide the user locally.

In this paper we present a system for non-visual exploration of virtual maps, based on haptic and auditory feedback. The system shares some ideas and goals with the works discussed above, yet it contains novel conceptual and technical features which will be illustrated in Section 2. Section 3 describes the design of two experiments aimed at assessing the validity of the proposed approach. Both involve non-visual exploration of purposedly simple virtual maps: at this stage of the work, the goal is not to test the system on real-word scenarios, but rather to assess the relative contributions of haptics and audition (and especially the added benefits of their integration) in controlled conditions with a small number of variables involved. Experimental results are presented and analyzed in Section 4, while Section 5 is devoted to global discussion and future developments.

[1] A 6-DOF position input/3-DOF force output haptic device with a stylus grip.

2. Auditory-haptic virtual maps

2.1. Concept and design

Our work focuses on the development of a system that helps a user to explore a non-visual map in order to acquire, organize, and revise knowledge about an unknown or unfamiliar environment before actually visiting it. As already discussed, it has been previously shown that users who are given information and mapping of an unknown space before arriving to it improve substantially their O&M performance in the real environment (Lahav and Mioduser, 2008; Jeon et al., 2012). This kind of assistive technology can be helpful especially for the acquisition of knowledge in indoor and domestic environments, with implications for safety, and recognition of furniture and home appliances.

With respect to previous works, the main distinguishing features of our approach can be summarized as follows.

First, the main driving concept in the use of auditory and haptic feedback is the exploitation of the complementarity between the two sensory modalities. Haptics, as a proximal sense, should provide information about local features in the virtual map being explored, whereas audition, as a distal sense, should provide global information about user's position within the map. In light of our initial discussion, this also means that haptic feedback mostly supports spatial cognition mechanisms based on path-integration, whereas auditory feedback mostly supports mechanisms based on global landmarks (Loomis et al., 2001).

Second, the above requirements are fulfilled through a design that employs principles of ecological and minimalistic rendering, i.e. a design that uses intuitive metaphors and provides a limited amount of information in order to avoid cognitive overload. Haptic feedback is based on a tactile bas-relief metaphor (ecological rendering), in which only the height profile (minimalistic rendering) of a 3D map is sensed by the user through his/her finger. 3D auditory feedback is based on the already mentioned "ears in hand" metaphor (ecological rendering) (Magnusson et al., 2006), in which only an abstract anchor sound at the center of the map is simulated (minimalistic rendering).

Third, the implementation of this design uses a novel hardware and software system. Specifically, it takes the form of an "audio-tactile tablet" in which active exploration through a tactile mouse allows to combine haptic and auditory information on a working area the size of a tablet, which makes the system low-cost and portable. Even more importantly, the implementation renders 3D sound binaurally through headphones, using an engine developed by some of the present authors (Spagnol et al., 2013; Geronazzo et al., 2014). The engine is based on head-related transfer functions (HRTFs), i.e. the frequency- and location-dependent acoustic transfer functions between the sound source and the eardrum of a listener. If a set of HRTFs is available (one pair of left and right HRTFs for each desired sound source position), an anechoic sound source can be virtually located in space by convolving it with the corresponding pair of left and right HRTFs and presenting the resulting binaural signal at the listener ears through headphones (see e.g. Xie, 2013 for an overview). As detailed next, in this work binaural rendering is adapted to individual anthropometry of the user, with improved perceptual results (Geronazzo et al., 2014).

2.2. Binaural 3D audio rendering

Since the recording of individual HRTFs is both time- and resource-consuming, alternative techniques for estimating them in different and more convenient ways are highly desirable. A common but nonoptimal choice is to employ a pre-defined HRTF set (e.g., recorded on a dummy head built according to average anthropometric data, such as the KEMAR mannequin (Gardner and Martin, 1995)) for any possible listener. However, individual anthropometric features of the human body shape the HRTFs and affect heavily the perception and the quality of the rendering (Møller et al., 1996): accordingly, there is a growing interest in more advanced HRTF selection techniques, that provide a user with his/her "best matching" HRTF set extracted from a public database, based on objective or subjective criteria (Seeber and Fastl, 2003; Zotkin et al., 2004; Katz and Parseihian, 2012).

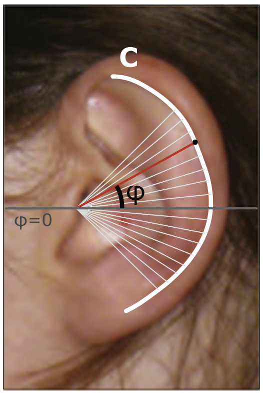

Fig. 1. 3D sound rendering: Estimation of a subject's pinna contours and distances from the ear canal entrance, from which the mismatch function with HRTF sets in the database are computed.

Our system employs a novel HRTF selection technique developed by some of the present authors (Geronazzo et al., 2014) and briefly summarized here. For a given subject, a best matching HRTF set is selected from a database based on relevant individual anthropometric features. Specifically, one image of the subject's pinna is used to compute an *ad hoc* defined mismatch function between the main pinna contours and corresponding spectral features (frequency notches) of the HRTFs in the database. The mismatch function is defined according to a ray-tracing interpretation of notch generation (Spagnol et al., 2013): as shown in Fig. 1, the main notches of a HRTF can be extracted with reasonable accuracy by calculating the distances between a point located approximately at the ear canal entrance and the points at the border of the helix (the contour labeled as C in Fig. 1), which is thought to be responsible for the main pinna reflections.

For a given elevation ϕ of the incoming sound, the distance between the pinna reflection point and the ear canal entrance is $d(\phi) = ct(\phi)$, where $t(\phi)$ is the temporal delay between the direct and reflected rays and c is the speed of sound. Assuming each reflection to be negative and responsible for a single notch (Spagnol et al., 2013), the corresponding notch frequency, $f_0(\phi)$, is estimated as

$$f_0(\phi) = \frac{c}{2d_c(\phi)}. \tag{1}$$

Given a subject whose individual HRTFs are not available, the mismatch m between his/her f_0 notch frequencies (estimated from Eq (1)) and the notch frequencies F_0 of an HRTF set in the database is defined as

$$m = \frac{1}{|\phi|} \sum_{\phi} \frac{|f_0(\phi) - F_0(\phi)|}{F_0(\phi)}, \tag{2}$$

where elevation ϕ spans all the available frontal angles in the database. The HRTF set that minimizes m is then selected as the best HRTF set in the database for that subject.

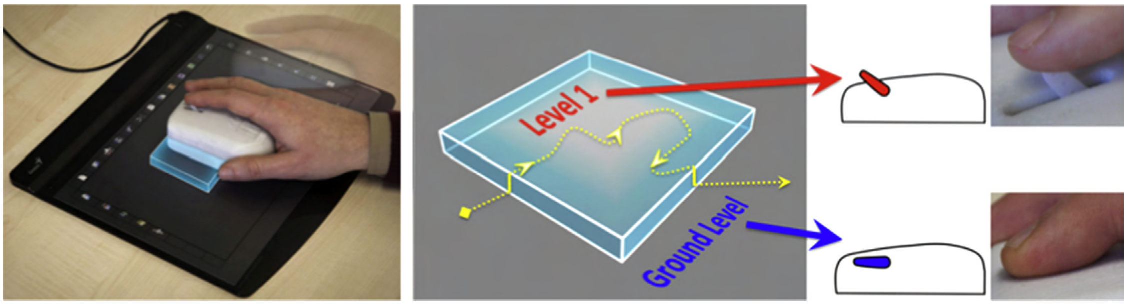

Fig. 2. Tactile rendering of a virtual object with the TAMO, in which the height of the mouse lever (in contact with the finger) simulates object height.

2.3. Tactile rendering

Haptic rendering is provided through the TActile MOuse (*TAMO*) device, recently developed by some of the present authors (Brayda et al., 2013). This device provides tactile stimuli using a mouse-like interaction metaphor, as depicted in Fig. 2.

The TAMO moves on a sensing tablet (in this case having size 210 × 297 mm^2) which represents the surface on which the virtual map is displayed. The TAMO renders haptically the presence of virtual objects and surfaces by moving a lever, placed above the device, and creating a tactile contact on the subject's fingertip. As shown in Fig. 2, when the mouse reaches a virtual object on the map the stepper signals a virtual edge of a given height. The tactile feedback corresponding to the minimum horizontal displacement in the workspace is called *taxel*: information related to a single tactile unit are stored and rendered like pixels in vision (Siegel, 2002). TAMO generates a taxel for each pixel of the working area, similarly to a tactile bas-relief representation. The lever moves from the ground horizontal position $\varphi_0 \approx 0^\circ$ to a nearly vertical position, corresponding to a rotation of $\varphi_{max} \approx 80^\circ$. In particular, the TAMO hosts a servomotor which is position-controlled and has a torque of 0.12 Nm, which is therefore transferred to the fingertip. The minimum lever rotation (0.7°) can be achieved by the motor in 2 ms. The lever is 15 mm long; its rounded cap has a 1.6 mm radius and is 1.5 mm wide. Participants were not forced to use the lever in a mandatory way: this was done for two reasons: first, forcing participants to adopt a fixed finger posture may distract him/her from the task; second, we observed in past experiments that some participants were touching the side of the lever to determine its distance from the upper shell of the TAMO, therefore explicitly touching the lever all along its length. Others, preferred to let the fingertip be lifted by the lever cap.

In addition to tactile feedback, the mouse-like metaphor employed in the TAMO also provides users with proprioceptive feedback: during exploration of the working area, users perceive the position of their hands, e.g. relative to the starting position of the exploration task. In this sense proprioception provides users with knowledge about their position on the working area through path-integration mechanisms.

While most haptic devices (e.g, the Phantom) are expensive and delicate hardware, the TAMO provides a low-budget, portable, and robust solution for specific application scenarios. Previous experimental result has shown that the combination of tactile feedback and active exploration appears adequate for the recognition and reconstruction of simple geometries, and, additionally, that the ecological and minimalistic feedback guarantees a rapid learning curve (Brayda et al., 2013).

2.4. System architecture and implementation

Fig. 3 shows a schematic view of the overall system architecture. All the experimental conditions are managed in Matlab. Information about the current status (2D position on the tablet) of the TAMO is also managed in Matlab, and is used to drive both the tactile and the 3D auditory rendering. The latter is realized in Pure Data,[2] an open source real-time environment for audio processing. Communication is managed through the Open Sound Control (OSC) protocol. Tactile feedback is rendered through the TAMO lever, while auditory feedback is rendered through headphones.

The implementation was tested to verify that real-time constraints are satisfied and that the auditory and haptic stimuli are synchronized within a coherent perceptual integration time window. We measured the latency between the two stimuli as follows: two condenser microphones connected to an audio card working at 192 kHz sampling rate were placed at the headphones coupler and near the lever, and latency was estimated as the time between the activation of the TAMO lever (detected by means of the noise of the TAMO mechanical engine) and the audio output at the earphones. Based on these measurements we chose a refresh rate of 80 ms, which is larger than the measured haptic-audio delay (68 ms) and therefore guarantees a consistent refresh. At the same time, it guarantees continuous integration of multimodal perception because neural information from different senses occur at approximately the same time, thus being associated with the same physical event (Holmes and Spence, 2005). Since signals coming from different sensory modalities have different time-of-arrivals and processing time in given brain area, a temporal window of about 200 ms ensures multisensory integration.

3D audio in Pure Data is rendered using the selection procedure described in Section 2.2. To this end, the acquisition of one pinna image of each system user is required in order to compute the mismatch between his/her manually traced contours and notch central frequencies. The chosen HRTF database is the CIPIC (Algazi et al., 2001), which contains HRTF sets measured in the far field (i.e., no distance information is available in these far-field compensated HRTFs) for 45 subjects, with azimuth angles spanning the range [0°,360°) and elevation [−45°,230.625°].

A virtual *anchor sound* is placed at the center of the map, and is spatially rendered according to the position of the user (mouse) relative to the center. The auditory stimulus is a continuous train of repeated 40 ms gaussian noise bursts with 30 ms of silence between each burst. Similar types of stimuli have been previously employed and found to be effective in localization tasks (Katz et al., 2012b), in alternative to single white noise bursts (Walker and Lindsay, 2006). The maximum amplitude of the raw stimulus at the entrance of the ear canal is set to 60 dB(A), and users can then manually adjust this default value in order to obtain a subjectively comfortable level. Since the HRTF based rendering only provides the angular position (azimuth and elevation) of the anchor sound, distance is rendered through an inverse square law of sound attenuation. A 25 px circular neighborhood is defined around the anchor sound, in which the sound pressure level remains constant at its maximum value, and a central *inside-the-*

[2] http://puredata.info

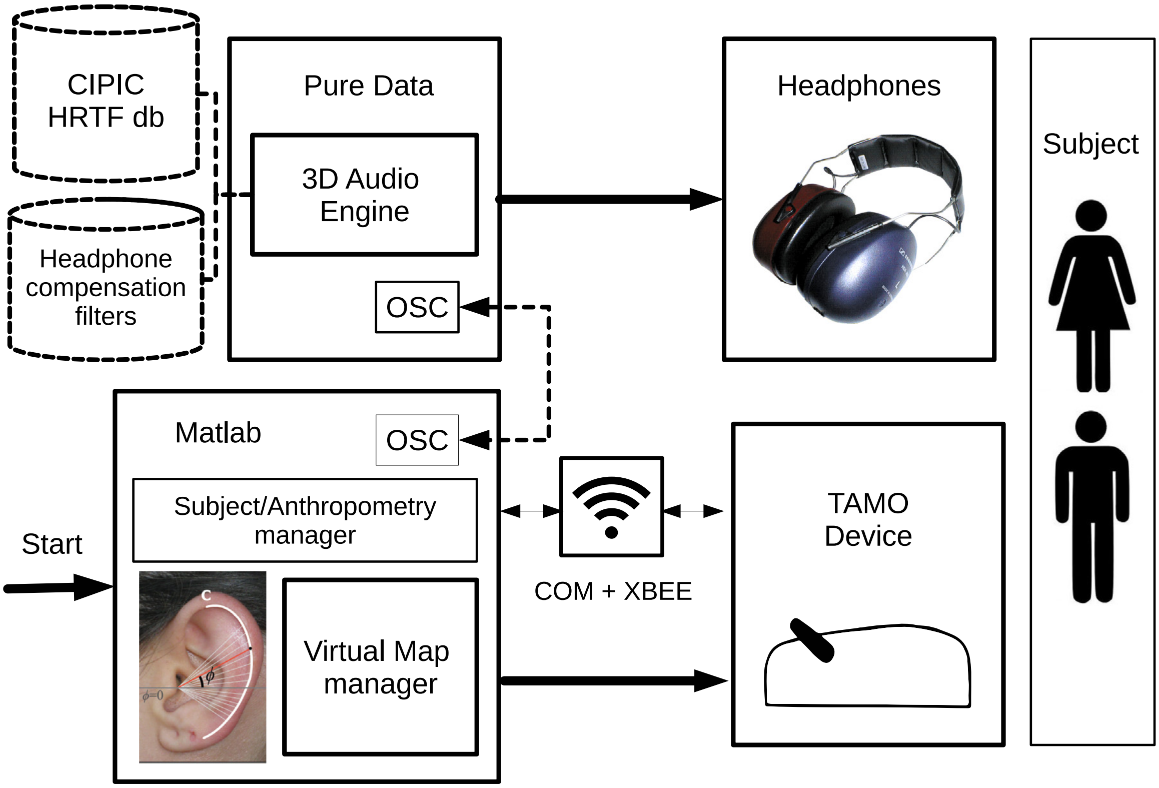

Fig. 3. A schematic view of the system architecture.

head perception is produced through a static filtering with center-front HRTFs (i.e., left/right HRTF where azimuth and elevation are equal to 0°) (Toole, 1970).

Audio output is managed through a Roland Edirol AudioCapture UA-101 board working at 44.1 kHz sampling rate, and delivered to Sennheiser HDA 200 headphones, which provide effective passive ambient noise attenuation. Additionally, their frequency response has no major peaks or notches between in the range 0.1–10 kHz (that would corrupt localization cues in the HRTF (Masiero and Fels, 2011)) and is almost independent on re-positionings on the users' head (Geronazzo, 2014). Nonetheless, the spatialized auditory stimuli are also processed through an equalization filter that compensates the non-flat headphone response, previously measured on a dummy head without pinnae: this non-individual compensation guarantees effective equalization of the headphones up to 8–10 kHz on average and simulated a realistic application scenario where it is not always feasible to design individualized headphone compensation filters (Geronazzo, 2014).

Boundaries of the virtual map on the working area are signalled both haptically and auditorily. Specifically, virtual "walls" at the boundaries are rendered by setting the TAMO lever to its maximum rotation φ_{max}. When the device moves beyond the walls the lever moves alternatively from φ_{max} to $\varphi_{max} - 30°$ at refresh rate, while the auditory feedback is stopped.

3. Experimental design

The relative contribution of haptic and auditory feedback, as well as their cross-modal effects, was evaluated by means of two experiments involving the exploration of simple virtual spatial maps.

Both the experiments define very simple maps, that are still far from real-world scenarios. It has to be emphasized that the goal of this work is to assess the relative contributions of the two modalities, and especially the added benefits of augmenting the virtual map with anchor sounds that complement local haptic information with global auditory information. Therefore, in order to limit the number of variables involved in such an assessment, very simple scenarios are explored at this stage of the work.

3.1. Subjects and apparatus

Eleven subjects (7 males and 4 females whose age varied from 21 to 40 with mean 28.2, SD 5.5) participated in the experiments. All subjects reported normal hearing after estimation of their hearing thresholds through an adaptive maximum likelihood procedure (Green, 1993). Only sighted users were involved in these experiments. This choice is in line with other previous studies (Walker and Lindsay, 2006) and is supported by the literature discussed in the introduction, which suggests that spatial representations are partially amodal (Loomis et al., 2002; Klatzky et al., 2003), and reports similarities between learning modalities and between groups of sighted, late-blind, and early-blind users (Afonso et al., 2010). More specifically, previous results by some of the present authors have provided evidence that visually impaired and sighted adults behave similarly in tactile map navigation using the TAMO (Brayda et al., 2015).

Both experiments were performed in a silent booth. All subjects were blindfolded and were only provided haptic and auditory feedback, through the TAMO and headphones, respectively, according to the specifications provided in Section 2.

3.2. Experiment #1: goal reaching

The first experiment has a preliminary nature and is not specifically focused on the acquisition of spatial knowledge by the subjects. Rather, its main goal is to assess the ecological validity of the proposed metaphors (the bass-relief tactile metaphor and the "ears in hand" auditory metaphor). One second goal is to analyze

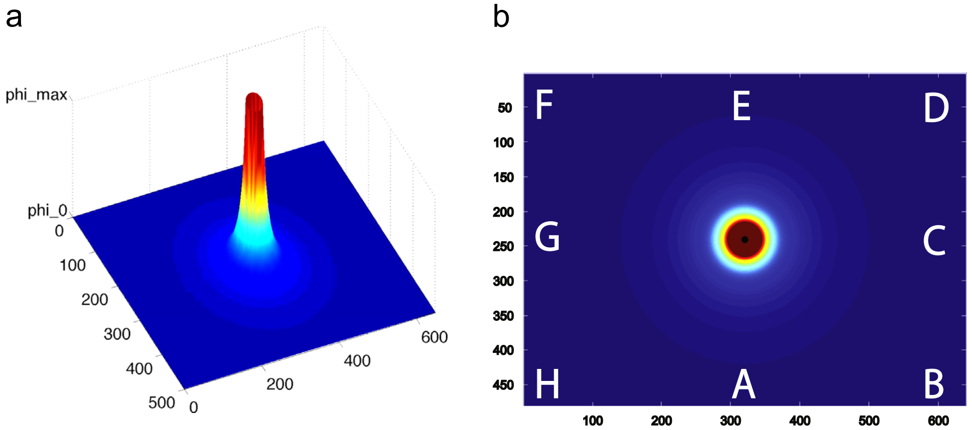

Fig. 4. Experiment #1: (a) The virtual maps. Haptic elevations are visualized through a height/color map, starting from blue (φ_0) and reaching till dark red (φ_{max}). (b) The goal is the central black dot. Starting positions for the audio-haptic exploration are marked in lexicographic order. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

the differences and the complementarity of auditory and tactile information, by means of behavioral and performance indicators collected from experimental subjects. In particular, the experiment allows to assess the effectiveness of the chosen approach with regard to auditory feedback, i.e. the use of an anchor sound signalling the center of the virtual map.

Such assessment is obtained through a goal reaching task, in which experimental subjects have to reach the center of the map under different (unimodal and bimodal) feedback conditions. Using Wiener's terminology (Wiener et al., 2009), assessing users' ability to locate and reach a target point in the map has relevant implications for the first level of spatial knowledge (i.e., knowledge about a point in space).

3.2.1. *Stimuli*

Anchor sounds were rendered according to the approach described in Section 2.2. Specifically, a spatialized sound source, represented by the black dot in Fig. 4, was placed at the center of the map. No height information was provided in the auditory feedback, i.e. only 2D information was mapped to azimuth and distance rendering.

On the other hand, the center of the map was rendered haptically through a height profile in which height varies as the inverse square of the distance from the center (see Fig. 4). This inverse square dependence was parametrized in such a way that the most distant points from the center (points B, D, F, H in Fig. 4(b)) correspond to the minimum rotation φ_0 of the TAMO lever, while all the points inside a 25 px neighborhood of map center correspond to the maximum rotation φ_{max}.

The choice of this particular height profile is somewhat arbitrary with respect to other possible options. A more rigorous choice (i.e. one that respects the proximal nature of haptic information in communicating the position of a point or small area on the map) would be a discrete profile in which the center is rendered as a small pillar in an otherwise flat map. Instead, the chosen profile provides a higher amount of information, as it allows the user to locally perceive a height gradient throughout the entire workspace, although the amount of haptic information provided at the map boundaries is less salient than that provided close to the center.

Three feedback conditions were provided:

- TAMO: unimodal haptic condition;
- 2D audio: unimodal auditory condition;
- TAMO + 2D audio: bimodal condition.

In order to minimize memory effects due to proprioception, the starting position had to be varied as much as possible between trials. To this end, subjects were asked to complete the task starting from eight different positions at the boundary of the workspace, as depicted in Fig. 4(b). Each of the three above feedback conditions was then repeated 8 times (one for each starting position) leading to a total of 24 trials per subject. The starting position was randomly varied according to a latin square order to minimize its effect on subjects' performance. Moreover, the three feedback conditions were randomized across trials, to average out learning effects.

3.2.2. *Procedure*

A brief tutorial session introduced the experiment. Subjects were verbally informed that they had to explore a virtual map using tactile and/or auditory information, and the exploration metaphors for the two modalities were described. Subjects were then instructed about the nature and meaning of the haptic and auditory stimuli, and were informed that their goal was to reach the center of the map as quickly as possible, using either auditory or tactile information, or both. During the task subjects were blindfolded: they were initially guided by the experimenter to the experimental location, and at each trial their hand was guided by the experimenter at the corresponding starting position. Finally, subjects were instructed to maintain a fixed head orientation, in order to be consistent with the “ears in hand” metaphor for auditory rendering.

Each trial was completed if the subject entered a 25 px neighborhood defined around the map center, and remained inside this area for 1.2 s. For each trial, the subject's trajectory in time was saved for further analysis.

3.3. *Experiment #2: cognitive map construction*

The second experiment is aimed at assessing whether the proposed approach and setup provides subjects with spatial knowledge about a simple virtual environment (or “survey knowledge”, using Wiener's terminology (Wiener et al., 2009)). This assessment is carried out by investigating to what extent subjects navigating freely in the environment are able to construct coherent spatial cognitive maps, in which the mental representation preserves the main topological and metric properties.

To this end, the experiment uses the simplest possible map, in which the position and the size of a single cube inside the

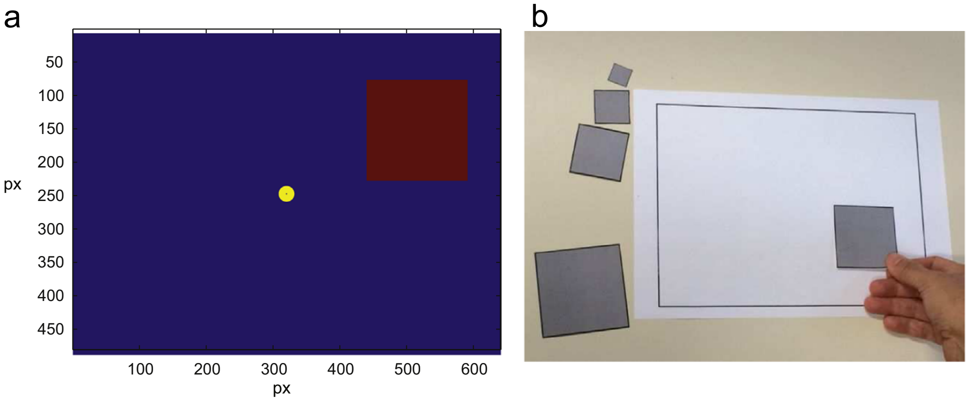

Fig. 5. Experiment #2: (a) Example of a map used in a trial, where the yellow dot and the red square represent the anchor sound and a virtual cube, respectively; (b) Example of a subject reconstructing an explored map, by choosing one among a set of five cardboard squares and placing it on a 210 × 297 mm a paper sheet. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

workspace have to be learned (see Fig. 5(a) for an exemplary map). Such a simple map is defined through a minimal number of variables. As such it serves to the scope of this paper as it allows to define simple and non-ambiguous performance metrics to quantify the relative contributions of tactile and auditory feedback, as well as their cross-modal effects, in affecting the accuracy of the constructed cognitive maps.

3.3.1. *Stimuli*

Each trial consisted in a virtual map in which a cube was rendered haptically. The horizontal coordinates of the center of mass of the cube were randomly generated at each trial, with the additional constraints that the base of the cube could not cross the boundaries of the workspace and could not include a neighborhood of the center with radius equal to the cube edge.[3] Two different cube sizes, small edge (80 px) and large edge (150 px), were considered. Moreover, the following three feedback conditions were used:

- TAMO: unimodal haptic condition;
- TAMO + 2D audio: bimodal condition with haptic feedback and 2D anchor sound;
- TAMO + 3D audio: bimodal condition with haptic feedback and 3D anchor sound.

In the second condition, the anchor sound at the center of the map was rendered in the same way as in Experiment #1. In the third condition, the anchor sound was rendered according to 3D coordinates, so that when the mouse was on top of the virtual cube the relative elevation between sound source and listener changed accordingly. Cube sizes were presented with two repetitions, yielding a total of $2 \times 3 \times 2 = 12$ stimuli arranged in latin square order. Both the cube size and the feedback condition were randomized.

3.3.2. *Procedure*

A brief tutorial session introduced the experiment. Subjects were already informed about the nature of tactile and auditory information and the exploration metaphors, based upon their experiences in the previous experiment. Subjects were then informed about the presence of a single virtual cube on the map, and that their goal was to find the cube and to estimate its size and location. They were also informed that the cube size could randomly change across trials. Subjects had one minute to explore the map.

After each trial, the subjects' ability to form a cognitive map of the explored environment was assessed as follows. They were asked to choose one among a set of cardboard squares, and to place it at their estimated location on a 210 × 297 mm area of a paper sheet. Although only two cube sizes were used in the stimuli, when giving their estimates subjects had to choose between the following five square sizes: 45 px, 80 px, 115 px, 150 px, 185 px (see Fig. 5(b)). In this way each of the two sizes actually used in the stimuli could potentially be confused with both a smaller and a larger one. Also, the physical size of the reconstructed map was exactly the same of the virtual map. We chose to discretize the size scale to have a convenient and non-arbitrary way of labeling correct answers.

This approach shares similarities with the one used in a recent related work (Picinali et al., 2014). There, subjects explored an environment, either through *in situ* real displacement or through active navigation in a virtual architecture. Afterwards, their ability to construct coherent mental maps was assessed by asking them to physically reconstruct the environment using LEGO® bricks.

The paper sheet was placed 50 cm away from the sensing tablet, and oriented orthogonally, in order to force subjects to keep the cognitive map in memory while moving from the experimental sitting position. For each trial, two parameters were saved from subjects' responses for further analysis: the size of the chosen cube, and its position of the reconstructed map.

Moreover, an informal questionnaire was performed at the end of the experimental session in order to collect subjective data on the interaction between participants and the system.

4. Experimental results

One of the 11 subjects involved in the two experiments was found to perform extremely poor in localizing virtual sounds, despite having reported normal hearing according to the procedure described in Section 3.1. In fact the very same subject had been excluded from the analysis of results in a previously published sound localization experiment (Geronazzo et al., 2014), in which his/her judgements were close to chance performance. Accordingly, this subject was discarded from the analysis also for the present experiments.

4.1. *Experiment #1 – results*

Experimental results were evaluated in terms of two main performance indicators for each trial: absolute reaching time and total traveled distance (i.e. the length of the trial trajectory).

[3] This ensures that minimum rendered elevation is inside the CIPIC range.

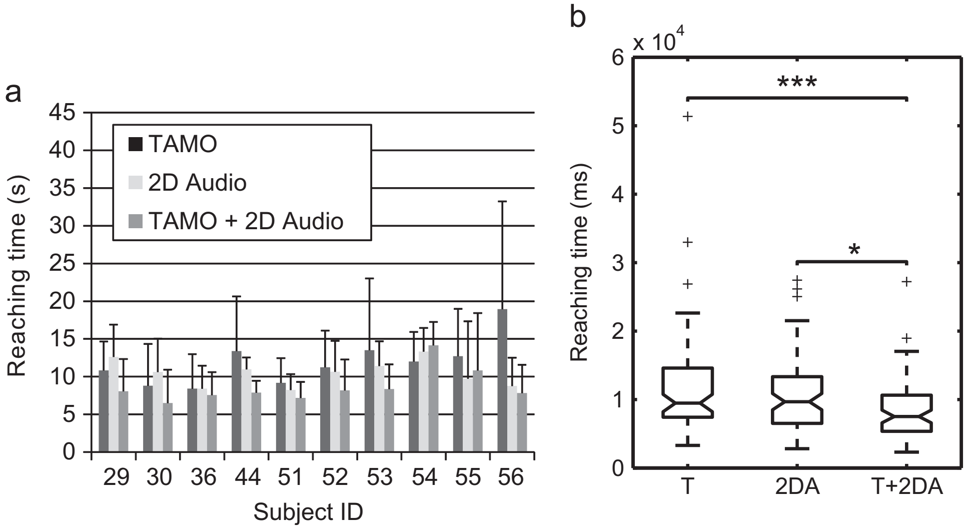

Fig. 6. Results for Experiment #1; (a) average and standard deviation (across all trials for each condition) of reaching times for each subject, and (b) global statistics on reaching times grouped by feedback condition. Asterisks and bars indicate, where present, a significant difference ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ at *post-hoc* test).

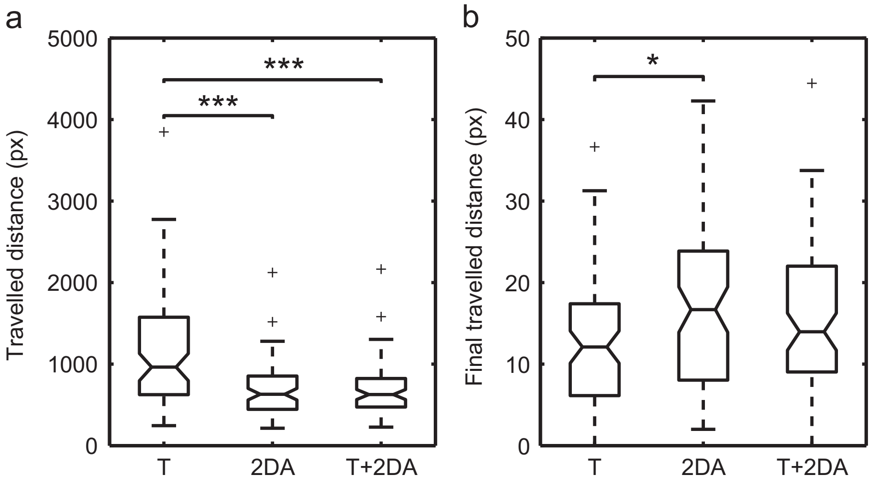

Fig. 7. Results for Experiment #1; (a) global statistics on total traveled distance, grouped by feedback condition, and (b) global statistics on “final” traveled distance, grouped by feedback condition. Asterisks and bars indicate, where present, a significant difference ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ at *post-hoc* test).

Previous literature has used other kinematic indicators of performance in similar tasks, using the segment between the starting point and the target as a reference trajectory. These includes such indicators as the cumulated error (the area of the surface between the trajectory and the reference Magenes et al. (1992)), the number of corrections (the number of variations in the trajectory direction that reduce the distance with the reference Magenes et al. (1992)), the aspect ratio (the maximum lateral deviation from the reference, divided by the reference length (Danziger and Mussa-Ivaldi, 2012)), and so on. However these indicators are mainly meaningful when the trajectory is reasonably close to the reference, i.e. to rectilinear motion, whereas in the present experiment trajectories had a much greater variability, and the total traveled distance is assumed to be a more salient indicator.

Fig. 6(a) shows average performances in terms of reaching times for each subject. The average improvement between the unimodal haptic condition and the bimodal condition was 25% ± 14%, with a maximum improvement of 59% and a minimum of 11%. The average improvement between unimodal auditory condition and the bimodal condition was 17% ± 14%.A Kruskal Wallis nonparametric one-way ANOVA with three levels of feedback condition was performed to asses the statistical significance of this indicator [$\chi^2(2) = 11.19$, $p < 0.05$]. Two *post-hoc* Wilcoxon tests (Fig. 6(b)) revealed statistically significant improvements in performance (decreases in reaching times) between the haptic unimodal condition and the bimodal condition ($p < 0.001$), and between the auditory unimodal condition and the bimodal condition ($p < 0.05$).

The global statistics for the total traveled distance is summarized in Fig. 7(a). A Kruskal Wallis nonparametric one-way ANOVA with three levels of feedback condition was performed [$\chi^2(2) = 31.84$, $p < 0.0001$]. Two *post-hoc* Wilcoxon tests revealed statistically significant improvements in performance (decreases of total traveled distances) between the haptic unimodal condition and the auditory unimodal conditions ($p < 0.0001$), and between the haptic unimodal condition and the bimodal condition ($p < 0.0001$). A further analysis on the traveled distance was performed in order to assess how users' performance changed in the vicinity of the target, where (i) proximal haptic information becomes more and more salient, and (ii) distal auditory information becomes on the contrary more confusing as the user approaches the anchor sound. This assessment was performed by analyzing the “final” traveled distance, i.e. the traveled distance in the last second of each trial. Fig. 7(b) shows the global statistics of such final traveled distance, subjected to a Kruskal Wallis non-parametric one-way ANOVA with three levels of feedback condition [$\chi^2(2) = 8.\ 19$, $p < 0.05$]. A *post-hoc* Wilcoxon test revealed a statistically significant deterioration in performance between the haptic unimodal condition and the auditory unimodal condition ($p = 0.004 < 0.05$). On the other hand, no statistically significant differences in performance were observed between the haptic unimodal condition and the bimodal condition ($p = 0.063$). The analysis on the final traveled distance also revealed that in 11 trials (on a total of 264) subjects were not moving in the last second (i.e. the final traveled distance was exactly 0): 7 of these trials belonged to the unimodal haptic condition, 4 belonged to the bimodal condition, and none belonged to the unimodal auditory condition, suggesting that haptic feedback provided users with non-ambiguous information about their completion of the task, while this was not the case for the unimodal auditory condition.

Taken together, these results confirm that effective multimodal (haptic-auditory) integration was achieved in the task. Analysis on reaching times and total traveled distance revealed that subjects adopted two different exploration strategies in the two unimodal conditions. Specifically, in the unimodal haptic condition they made on average faster movements with long trajectories, while in the unimodal auditory condition they followed shorter trajectories at a slower speed. The bimodal condition exhibits statistically significant improved reaching times and total traveled distances, showing that the two strategies are successfully integrated into exploration paths involving short trajectories traveled at fast speed.

Moreover, analysis on the final traveled distances shows that haptic information was more reliable than auditory information in the vicinity of the target. In particular, the smaller final traveled distances in the unimodal haptic condition show that on average subjects were more confident about having completed the task. On the contrary, the larger final traveled distances in the unimodal auditory condition show that in this case subjects had less clear information about the target. Nonetheless, it has also been shown that adding auditory feedback to haptic feedback was not detrimental in the vicinity of the target, as shown by the absence of statistically significant differences between the unimodal haptic condition and the bimodal condition.

4.2. Experiment #2 – results

Two indicators were extracted from the maps reconstructed by subjects after each trial (with cardboard squares on a 210 × 297 mm paper sheet): the size of the chosen square, and the position error, i.e. the absolute distance in the horizontal plane between the centers of mass of the square in the reconstructed map and the one in the virtual map. Together, these two indicators provide a measure of subjects' ability to construct coherent spatial cognitive maps preserving the main topological and metric properties of the explored virtual map. Fig. 8(a) and (b) shows two example trials, specifically the best trials of Subject 30: in both trials location errors were very small, and the estimated sizes were correct.

Fig. 9(a) shows individual results for the position error. It can be noticed that estimates in the two bimodal conditions (with 2D and 3D anchor sound, respectively) improved those in the unimodal haptic condition for seven subjects. The average improvement between the unimodal haptic condition and the bimodal condition with 2D anchor sound was 9.9%, and it raised to 16.6% for the bimodal condition with 3D anchor sound. A Kruskal Wallis non-parametric one-way ANOVA with three levels of feedback condition was performed to assess the statistical significance of this indicator [$\chi^2(2) = 6.92$, $p < 0.05$]. A *post-hoc* Wilcoxon test (Fig. 9 (b)) revealed a statistically significant decrease ($p < 0.05$) of position errors between the unimodal haptic condition and the bimodal condition with 3D anchor sound.

Fig. 10 visualizes the numbers of incorrect/correct estimates by all subjects (120 estimates in total), grouped by feedback condition. Again, it can be noticed that estimates in the two bimodal conditions improved with respect to the unimodal haptic condition. While only a slight improvement was observed in the first bimodal condition (2D anchor sound), a much stronger improvement was obtained in the second bimodal condition (3D anchor sound), which is also the only condition where the number of correct estimates exceeded incorrect ones. Results were analyzed through a Pearson's Chi-Square test of independence of incorrect/correct answers on feedback conditions. The effect of feedback condition was found to be significant [$\chi^2 = 7.\ 11$, $p < 0.05$].

An overall tendency to underestimate sizes was observed: the smallest proposed size (45 px) was chosen 29 times while the largest (185 px) was chosen 8 times on a total of 120 trials for all subjects. This effect can be related to the presence of a scaling factor in subject spatial representation with the TAMO, which was observed in previously published results (Brayda et al., 2010).

Taken together, results from Experiment #2 show that, compared to unimodal haptic feedback, bimodal feedback improved significantly the amount of spatial knowledge acquired by subjects while exploring the virtual maps, and their ability to reconstruct them accurately. The effect was especially large for the second bimodal condition, i.e. when 3D anchor sounds were used. In this case, improvements in both object location errors and estimation of object sizes were statistically significant.

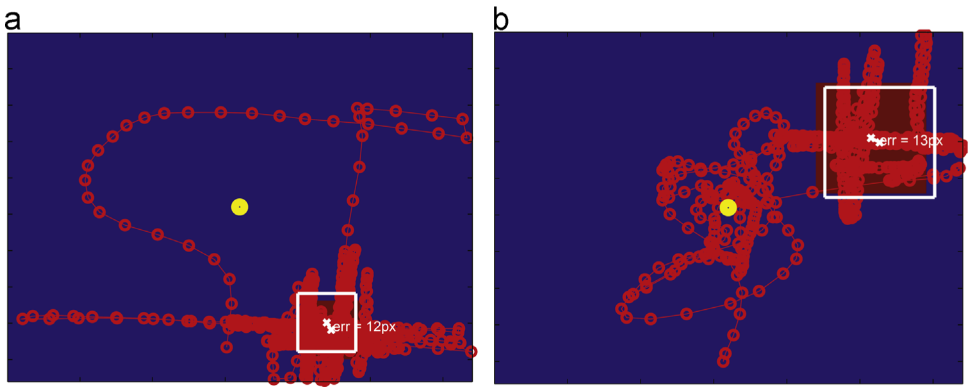

Fig. 8. Experiment #2. Two trials of Subject 30: (a) the cube with small (45 px) edge and (b) the cube with large (80 px) edge. Subject trajectories are shown in red. The object location errors in subject estimates (12 px and 13 px, respectively) are also shown. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

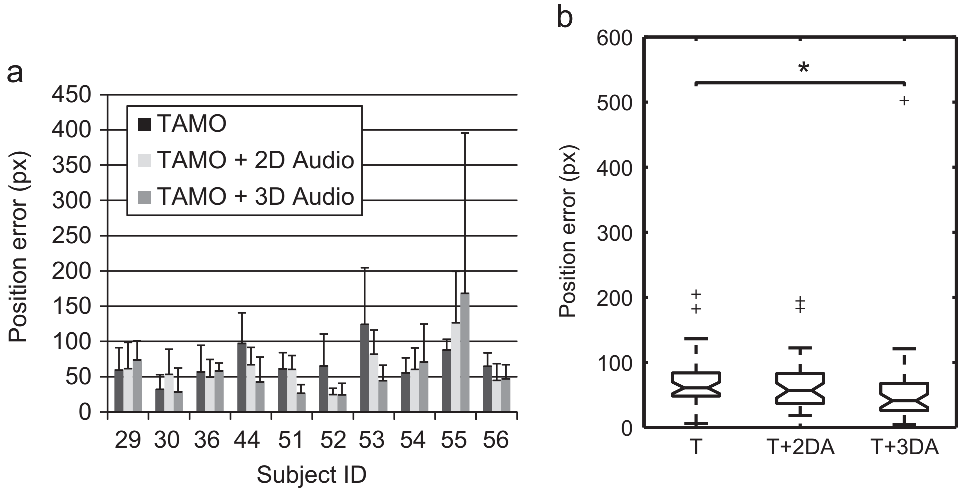

Fig. 9. Results for Experiment #2; (a) average and standard deviation (across all trials for each condition) of the position errors for each subject, and (b) global statistics on position errors grouped by feedback condition. Asterisks and bars indicate, where present, a significant difference ($^* p < 0.05$ at *post-hoc* test).

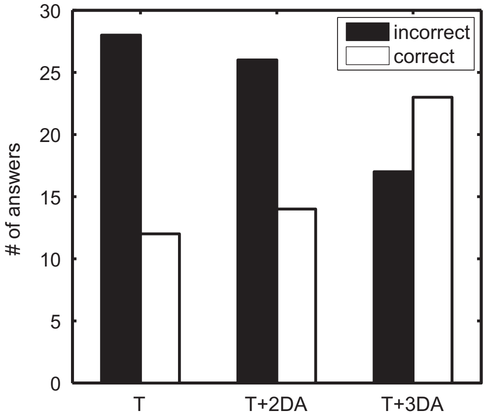

Fig. 10. Results for experiment #2: global count of correct/incorrect size estimations grouped by feedback condition.

5. General discussion and conclusions

The main goal of this work is to provide a first evaluation of a novel multimodal interactive system for non-visual (auditory-haptic) exploration of virtual maps. More precisely, the goal is twofold: first, we aim at validating the effectiveness and intuitiveness of the proposed haptic and auditory metaphors used to render the virtual maps; second, we aim at evaluating qualitatively and quantitatively the relative contributions of the two modalities in the exploration of virtual maps, in relation to subjects' performance in simple exploration tasks and subjects' ability to construct coherent cognitive representations of the explored maps.

Two experiments were designed to this end. Both use simple maps that are far from real-world scenarios, allowing to limit the number of variables involved in the definition of the maps, and to extract meaningful indicators of subjects' performance. The first preliminary experiment is not specifically focused on the acquisition of spatial knowledge: it analyzes the intuitiveness of the proposed metaphors, and the complementarity of auditory and tactile information, by means of behavioral and performance indicators in a goal reaching task. Instead the second experiment focuses on the acquisition of spatial knowledge of a simple virtual environment, and particularly on subjects' ability to form a coherent spatial cognitive map of a freely explored map in which the position and the size of a single object inside the workspace have to be learned.

The experimental data discussed in the previous section can be summarized into three main results. First, results from both experiments show that the system allowed subjects to effectively exploit the complementary natures of the "proximal" haptic modality and the "distal" auditory modality. This conclusion is supported by many observations. In Experiment #1, analysis on both reaching times and total traveled distances showed that subjects adopted two different exploration strategies in the two unimodal conditions, and that they successfully integrated them in the bimodal condition where exploration paths were traveled at faster speed. The only exception to this trend is the analysis of trajectory length, where statistically significant improvements were found between uni-modal haptic and bi-modal conditions but not between uni-modal auditory and bi-modal conditions. Moreover, analysis on the final traveled distances (i.e. trajectory lengths in the last second of each trial) showed that on average subjects reached the target more easily in the unimodal haptic condition than in the unimodal auditory condition, but it also showed that the addition of auditory feedback to haptic feedback was not detrimental to performance in the vicinity of the target. Results from Experiment #2 showed that the addition of spatialized anchor sound in the explored virtual maps improved significantly the amount of spatial knowledge acquired by subjects during exploration with respect to the unimodal haptic exploration condition. This conclusion is supported by statistical significance of the improvements in the accuracy of map reconstruction, measured in terms of object location errors and object size recognition.

Second, subjective data from Experiment #2 yielded a partially unexpected yet very interesting clear result. The improvements provided by 3D auditory feedback with respect to unimodal haptic exploration are strikingly larger than those provided by 2D auditory feedback. This finding confirms the importance of using individualized HRTF for binaural rendering through the approach described in Section 2.2, since individual characteristics of pinna shapes are particularly relevant for vertical localization of sound. It is worth mentioning that, despite the improved performance of 3D anchor sounds with respect to 2D anchor sounds, informal comments revealed that subjects did not consciously note the difference between the two auditory modalities.

Although, no formal evaluation was performed regarding the effectiveness and intuitiveness of the proposed minimalistic approach and the haptic/auditory rendering metaphors, informal observations and interviews with participants suggest that they were easily learned both in unimodal conditions and in bimodal condition. It has to be noted that subjects were only verbally informed about the exploration metaphors for the two modalities, but no training was provided before the experimental sessions.

Further investigations are needed in order to consolidate this result comparing generic HRTFs and several HRTF selection methods, such as ITD optimization procedures (Katz and Noisternig, 2014) and different anthropometric features (Middlebrooks, 1999). Each spatial audio technologies should be evaluated in both static localization tests and dynamic navigation tasks. A possible explanation for this result is that 3D anchor sounds provide intrinsically richer information than 2D anchor sounds, and specifically they transparently provide users with local information about being over the cube. The sudden change in elevation in proximity of the cube acts as an auditory trigger for navigation, while haptics plays the role of absolute local reference. We could have used a GUI to reproduce the position and the size of the virtual object. This would have made data collection easier. However, we chose to use a physical setup for two reasons: first, the physical setup has the same dimensions of the working space of the tablet, therefore on such setup the proprioceptive inputs learnt during the exploration of the virtual object can be used as reminiscent cues for the construction of the physical object. Second, such setup is very similar to what visually impaired individuals use when learning geometrical concepts. Our ultimate goal is in fact to compare our audio-haptic setup with state-of-art rehabilitation practices.

Admittedly, in the present work we did not include measures of cognitive load. However, we have previously and extensively studied the relation between the complexity of virtual objects delivered with the TActile MOuse and a measure of cognitive load (Brayda et al., 2013). In this work we deliver haptic stimuli that are similar to those eliciting, in our past works, low cognitive load values.

These outcomes provide a solid ground for the planning of further experiments in which (i) more complex scenarios can be created, and (ii) higher-level cognitive tasks can be tested. Regarding the first point, current work is directed at the rendering of maps with larger numbers of simple geometrical objects, and possibly anchor sounds with different locations in the map and different sound timbres (using a preliminary version of the proposed system, results have already been obtained regarding non-visual recognition of object shapes (Geronazzo et al., 2015a)). The subsequent step is to use maps of real environments, starting e.g. from existing CAD data of indoor locations such as daily living environments. Regarding the second point, higher-level tasks include in particular wayfinding tasks in which subjects need to define optimal paths between two points in a map. In the case of maps of real environments, it will be possible to assess subjective performances in real environments and to compare performances of groups previously trained with the system with respect to groups exploring the real environment with no previous training. Further experiments with also be required to involve both sighted and visually-impaired (late-blind and early-blind) subjects. We emphasize that, although in general studies on sighted subjects may not allow to draw inferences on a non-sighted population, we have recently compared the two populations, in another task, however using the same TActile MOuse (Brayda et al., 2015). We reported substantial similarities between sighted and non sighted adults, and differences were accounted for by factors such as different visual experience and gender rather than the status of just being blind.

These results also provide several indications for improving the design and the implementation of the proposed system. In particular, the system could benefit from a larger workspace. Moreover the orientation of the mouse (or the orientation of subject's head) on the horizontal plane can be tracked in order to improve the interactivity of the spatial sound rendering.

Multimodal virtual environments for spatial data sonification and navigation are expected to substantially benefit in terms of accessibility, scalability and deployment, from audio rendering framework in mobile devices and web platforms (Geronazzo et al., 2015b). Finally, this technologies can be developed and tested with different haptic/tactile devices. Available options range from "traditional" stylus based devices (such as the Phantom) to upcoming novel technologies that enable tactile feedback on screens such as the TeslaTouch (Xu et al., 2011), which would allow to implement the proposed design on mobile devices.

Acknowledgments

This work was partially funded by the research project "PADVA - Personal Auditory Displays for Virtual Acoustics" (no. CPDA135702) of the University of Padova.

Appendix A. Supplementary material

Supplementary data associated with this paper can be found in the online version at http://dx.doi.org/10.1016/j.ijhcs.2015.08.004.

References

Afonso, A., Blum, A., Katz, B., 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 Cognit. 38 (5), 591–604.

Algazi, V.R., Duda, R.O., Thompson D.M., Avendano C., 2001. The CIPIC HRTF database. In: Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, USA, pp. 1–4.

Bach-y Rita, P., 2004. Tactile sensory substitution studies. Ann. N.Y. Acad. Sci. 1013, 83–91, PMID: 15194608 URL⟨http://www.ncbi.nlm.nih.gov/pubmed/15194608⟩.

Brayda, L., Campus, C., Chellali, R., Rodriguez G., 2010. Objective evaluation of spatial information acquisition using a visuo-tactile sensory substitution device. In: Ge, S., Li, H., Cabibihan, J.-J., Tan, Y. (Eds.), Social Robotics, Lecture Notes in Computer Science, vol. 6414, Springer, Berlin/Heidelberg, pp. 315–324.

Brayda, L., Campus, C., Gori, M., 2013. Predicting successful tactile mapping of virtual objects. IEEE Trans. Haptics 6 (4), 473–483. http://dx.doi.org/10.1109/TOH.2013.49.

Brayda, L., Memeo, M., Lucagrossi, L., 2015. The importance of visual experience, gender and emotion in the assessment of an assistive tactile mouse. IEEE Trans. Haptics 99, 1. http://dx.doi.org/10.1109/TOH.2015.2426692 ⟨http://ieeexplore.ieee.org.biblio.iit.it/xpl/articleDetails.jsp?tp=&arnumber=7095578&queryText%3Dthe+importance+of+visual+experience+gender⟩.

Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., Pietrini, P., 2008. Imagery and spatial processes in blindness and visual impairment. Neurosci. Biobehav. Rev. 32 (8), 1346–1360. http://dx.doi.org/10.1016/j.neubiorev.2008.05.002.

Danziger, Z., Mussa-Ivaldi, F.A., 2012. The influence of visual motion on motor learning. J. Neurosci. 32 (29), 9859–9869.

Driver, J., Spence, C., 2000. Multisensory perception: beyond modularity and convergence. Curr. Biol. 10 (20), R731–R735.

Dubus, G., Bresin, R., 2013. A systematic review of mapping strategies for the sonification of physical quantities. PLoS One 8 (12), e82491. http://dx.doi.org/10.1371/journal.pone.0082491.

Ernst, M.O., Bülthoff, H.H., 2004. Merging the senses into a robust percept. Trends Cognit. Sci. 8 (4), 162–169. http://dx.doi.org/10.1016/j.tics.2004.02.002.

Gardner, W.G., Martin, K.D., 1995. HRTF measurements of a KEMAR. J. Acoust. Soc. Am. 97 (6), 3907–3908.

Geronazzo, M., 2014. Mixed Structural Models for 3D Audio in Virtual Environments (Ph.D. thesis). Information Engineering, Padova.

Geronazzo, M., Spagnol, S., Bedin, A., Avanzini F., 2014. Enhancing vertical localization with image-guided selection of non-individual head-related transfer functions. In: Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2014), Firenze, 2014, pp. 4496–4500.

Geronazzo, M., Bedin, A., Brayda, L., Avanzini, F., 2014. Multimodal exploration of virtual objects with a spatialized anchor sound. in: Proceedings of 55th

International Conference on Audio Engineering Society, Spatial Audio, Helsinki, 2014, pp. 1–8.

Geronazzo, M., Kleimola, J., Majdak, P., 2015. Personalization support for binaural headphone reproduction in web browsers. In: Proceedings of 1st Web Audio Conference, Paris, France, pp. 1–8.

Gibson, E.J., Pick, A.D., 2000. An Ecological Approach to Perceptual Learning and Development. Oxford University Press, New York.

Green, D.M., 1993. A maximum-likelihood method for estimating thresholds in a yes-no task. J. Acoust. Soc. Am. 93 (4), 2096–2105.

Holmes, N.P., Spence, C., 2005. Multisensory integration: space, time, & superadditivity. Curr. Biol. 15 (18), R762–R764. http://dx.doi.org/10.1016/j.cub.2005.08.058.

Jeon, M., Nazneen, N., Akanser O., Ayala-Acevedo, A., Walker, 2012. B., Listen2dRoom: helping blind individuals understand room layouts. In: Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'12), pp. 1577–1582.

Katz, B.F., Noisternig, M., 2014. A comparative study of interaural time delay estimation methods. J. Acoust. Soc. Am. 135 (6), 3530–3540 ⟨http://scitation.aip.org/content/asa/journal/jasa/135/6/10.1121/1.4875714⟩.

Katz, B.F.G., Parseihian, G., 2012. Perceptually based head-related transfer function database optimization. J. Acoust. Soc. Am. 131 (2), EL99–EL105.

Katz, B.F.G., Kammoun, S., Parseihian, G., Gutierrez, O., Brilhault, A., Auvray, M., Truillet, P., Denis, M., Thorpe, S., Jouffrais, C., 2012a. Navig: augmented reality guidance system for the visually impaired. Virtual Real. 16 (4), 253–269.

Katz, B.F.G., Kammoun, S., Parseihian, G., Gutierrez, O., Brilhault, A., Auvray, M., Truillet, P., Denis, M., Thorpe, S., Jouffrais, C., 2012b. NAVIG: augmented reality guidance system for the visually impaired: Combining object localization, GNSS, and spatial audio. Virtual Real. 16 (4), 253–269. http://dx.doi.org/10.1007/s10055-012-0213-6.

Klatzky, R.L., Lippa, Y., Loomis, J.M., Golledge, R.G., 2003. Encoding, learning, and spatial updating of multiple object locations specified by 3-d sound, spatial language, and vision. Exp. Brain Res. 149 (1), 48–61.

Lahav, O., Mioduser, D., 2004. Exploration of unknown spaces by people who are blind using a multi-sensory virtual environment. J. Special Educ. Technol. 19 (3), 15–23.

Lahav, O., Mioduser, D., 2008. Construction of cognitive maps of unknown spaces using a multi-sensory virtual environment for people who are blind. Comput. Human Behav. 24 (3), 1139–1155. http://dx.doi.org/10.1016/j.chb.2007.04.003.

Lahav, O., Schloerb, D.W., Kumar1, S., Srinivasan, M.A., 2012. A virtual environment for people who are blind—a usability study. J. Assist. Technol. 6 (1), 1–21.

Lecuyer, A., Mobuchon, P., Megard, C., Perret, J., Andriot, C., Colinot, J.-P., 2003. HOMERE: a multimodal system for visually impaired people to explore virtual environments, in: Virtual Reality, 2003. Proceedings. IEEE. Los Angeles, CA, USA, pp. 251 – 258. http://dx.doi.org/10.1109/VR.2003.1191147.

Loeliger, E., Stockman, T., 2014. Wayfinding without Visual Cues: Evaluation of an Interactive Audio Map System. Interacting with Computers 26, 403–416. http://dx.doi.org/10.1093/iwc/iwt042.

Loomis, J.M., Klatzki, R.L., Golledge, R.G., 2001. Navigating without vision: basic and applied research. Optom. Vis. Sci. 78 (5), 282–289.

Loomis, J.M., Lippa, Y., Klatzki, R.L., Golledge, R.G., 2002. Spatial updating of locations specified by 3-d sound and spatial language. J. Exp. Psychol. 28 (2), 335–345.

Magenes, G., Vercher, J.L., Gauthier, G.M., 1992. Hand movement strategies in telecontrolled motion along 2-d trajectories. IEEE Trans. Syst. Man Cybern. 22 (2), 242–257.

Magnusson, C., Danielsson, H., Rassmus-Gröhn K., 2006. Non visual haptic audio tools for virtual environments. In: McGookin, D., Brewster, S. (Eds.), Haptic and Audio Interaction Design, No. 4129 in Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 111–120.

Masiero, B., Fels, J., 2011. Perceptually robust headphone equalization for binaural reproduction. In: Audio Engineering Society Convention 130, pp. 1–7.

Menelas, B.-A.J., Picinali, L., Bourdot, P., Katz, B.F.G., 2014. Non-visual identification, localization, and selection of entities of interest in a 3d environment. J. Multimodal User Interfaces 8 (3), 243–256.

Middlebrooks, J.C., 1999. Individual differences in external-ear transfer functions reduced by scaling in frequency. J. Acoust. Soc. Am. 106 (3), 1480–1492. http://dx.doi.org/10.1121/1.427176.

Møller, H., Søorensen, M., Friis, J., Clemen, B., Hammershoi, D., 1996. Binaural technique: do we need individual recordings?. J. Audio Eng. Soc. 44 (6), 451–469.

O'Regan, J.K., Noë, A., 2001. A sensorimotor account of vision and visual consciousness. Behav. Brain Sci. 24 (5), 883–917.

Picinali, L., Afonso, A., Denis, M., Katz, B.F.G., 2014. Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge. Int. J. Hum.–Comput. Stud. 72 (4), 393–407. http://dx.doi.org/10.1016/j.ijhcs.2013.12.008.

Révész, G., 1950. Psychology and Art of the Blind. Longmans, Green and Co., London.

Schloerb, D.W., Lahav, O., Desloge, J.G., Srinivasan, M.A., 2010. BlindAid: Virtual environment system for self-reliant trip planning and orientation and mobility training, in: Haptics Symposium, 2010 IEEE. Waltham, MA, pp. 363 –370. http://dx.doi.org/10.1109/HAPTIC.2010.5444631.

Seeber, B.U., Fastl, H., 2003. Subjective selection of nonindividual head-related transfer functions. In: Proceedings of 2003 International Conference on Auditory Display (ICAD03), Boston, MA, USA, pp. 259–262.

Siegel, M., 2002. Tactile display development: the driving-force for tactile sensor development. In: IEEE International Workshop 2002 HAVE on Haptic Virtual Environments and their Applications, pp. 115–118. http://dx.doi.org/10.1109/HAVE.2002.1106924.

Spagnol, S., Geronazzo, M., Avanzini, F., 2013. On the relation between pinna reflection patterns and head-related transfer function features. IEEE Trans. Audio Speech Lang. Process. 21 (3), 508–519.

Spence, C., 2011. Crossmodal correspondences: a tutorial review. Atten. Percept. Psychophys. 73, 971–995.

Toole, F.E., 1970. In-head localization of acoustic images. J. Acoust. Soc. Am. 48 (4B), 943–949. http://dx.doi.org/10.1121/1.1912233.

Usoh, M., Arthur, K., Whitton, M.C., Bastos, R., Steed, A., Slater, M., Brooks, F.P., Jr., 1999. Walking > walking-in-place > flying, in virtual environments. In: Proceedings of ACM SIGGRAPH (SIGGRAPH99), Los Angeles, pp. 359–364.

Viaud-Delmon, I., Warusfel O., 2014. From ear to body: the auditory-motor loop in spatial cognition. Audit. Cognit. Neurosci. 8, 283. http://dx.doi.org/10.3389/fnins.2014.00283, URL⟨http://journal.frontiersin.org/article/10.3389/fnins.2014.00283/full⟩.

Vorländer, M., 2008. Auralization: Fundamentals of Acoustics, Modeling, Simulation Algorithms and Acoustic Virtual Reality, 1st edition. Springer-Verlag, Berlin.

Walker, B.N., Lindsay, J., 2006. Navigation performance with a virtual auditory display: effects of beacon sound, capture radius, and practice. Hum. Factors: J. Hum. Factors Ergon. Soc. 48 (2), 265–278. http://dx.doi.org/10.1518/001872006777724507.

Wiener, J.M., Büchner, S.J., Hölscher, C., 2009. Taxonomy of human wayfinding tasks: a knowledge-based approach. Spat. Cognit. Comput. 9 (2), 152–165. http://dx.doi.org/10.1080/13875860902906496.

Xie, B., 2013. Head-Related Transfer Function and Virtual Auditory Display. J. Ross Publishing, Plantatation, FL., pp. 363–370.

Xu, C., Israr A., Poupyrev I., Bau O., Harrison C., 2011. Tactile display for the visually impaired using TeslaTouch. In: CHI '11 Extended Abstracts on Human Factors in Computing Systems, CHI EA '11, ACM, New York, NY, USA, 2011, pp. 317–322. http://dx.doi.org/10.1145/1979742.1979705 URL ⟨http://doi.acm.org/10.1145/1979742.1979705⟩.

Zotkin, D., Duraiswami, R., Davis, L., 2004. Rendering localized spatial audio in a virtual auditory space. IEEE Trans. Multimed. 6 (4), 553–564. http://dx.doi.org/10.1109/TMM.2004.827516.