

Multimodal exploration of virtual objects with a spatialized anchor sound

Michele Geronazzo¹, Alberto Bedin¹, Luca Brayda², Federico Avanzini¹,

¹*Dip. Ingegneria dell'Informazione, University of Padova, Padova, Via Gradenigo 6/B, 35131, Italy*

²*Dep. of Robotics, Brain and Cognitive Sciences, Fondazione Istituto Italiano di Tecnologia, Genova, Via Morego 30, 16163, Italy*

Correspondence should be addressed to Michele Geronazzo (geronazzo@dei.unipd.it)

ABSTRACT

A multimodal interactive system for audio-haptic integration is presented in this paper. Preliminary subjective tests with a virtual reality setup were conducted with the goal of interpreting cognitive mechanisms and improving performances in orientation & mobility protocols for visually impaired subjects, where spatial representations need to be developed using residual sensory channels. An object recognition experiment was performed in order to investigate the contribution of dynamic spatial audio cues when integrated with haptic feedback. Audio cues took the form of anchor sound delivered through headphones using customized Head-related Transfer Functions (HRTFs). This setup was employed in the exploration of simplified virtual audio-tactile environments. Overall results on recognition time reveal a relationship between anchor position and object shape. Moreover, a qualitative analysis of the exploration paths highlights behavioral changes between unimodal and multimodal conditions.

1. INTRODUCTION

The concurrent presence of multiple sensory channels in multimodal virtual environments allows users to dynamically switch between modalities during the interaction. In particular, sensory augmentation through additional modalities and sensory substitution techniques [1] are compelling ingredients in presenting information non-visually, when visual bandwidth is overloaded, when data are visually occluded, or when the visual channel is not available to the user (e.g., for visually impaired people). Multimodal systems for representation of spatial information could largely benefit from audio engines that exploit known mechanisms of spatial hearing.

Walker et al. [2] analyzed the effects of dynamic non-speech auditory beacons in navigation using virtual auditory displays. Auralization through headphones accompanied user locomotion in searching tasks toward the beacon sound, and an overall increase in speed and accuracy was reported.

A 3D audio virtual environment was developed by Afonso et al. [3] aiming to investigate structural properties of spatial representation in people with different level of blindness. Accordingly, six anchor points (ecological sounds) were acoustically spatialized on the horizontal

plane around the subject which had to learn sounds position by locomotion.

In the above mentioned studies, the use of spatialized anchors was a compelling element in providing global and dynamic information on sound sources and their relationship with the surrounding environment, in a way that cannot be substituted by other modalities (i.e., visual or tactile). A similar approach is used in this work, with additional specific attention to the individual characteristics of the listener's body which acts as a filter on the incoming sound [4].

Specifically, our work focuses on how auditory and tactile modalities can be combined to improve spatial processing by means of a "virtual tactile tablet" [5]. The proposed setup includes a tactile display based on a touch-sensitive tablet which allows users to acquire haptic information for each interaction point in the working area, and a continuous global auditory reference. The main idea relies on rendering local information through haptics, i.e. impacts caused by body actions, while rendering global information through acoustics, i.e. constant references based on body position and orientation. Here a minimalist haptic feedback, the TACTile MOuse (TAMO) device [6] provides simple tactile stim-

uli through a mouse-like interactive device, which allows the construction of a cognitive map through touch; in addition, a 2D audio anchor in the middle of the map allows an exploration of virtual objects which benefits from a reference point. In this very first stage, we employ simple haptic 3D geometries, i.e. parallelepiped and cylinder, rather than complex shapes in order to limit information that may disrupt subject attention from the anchor point.

The paper is organized as follows. Section 2 outlines the motivations behind our multimodal interactive system. In Section 3 and 4, we discuss a preliminary object recognition experiment using either unimodal (haptic) or bimodal (haptic and auditory) feedback, and show that exploration strategies in the recognition task exhibit qualitative changes between the two conditions.

2. MOTIVATIONS

Sighted people rely on vision, as this is the predominant sensory modality to perform most of everyday activities. To this regard, tasks that look straightforward and natural for sighted people, e.g. exploring and recognizing the shape of an object, become challenging for blind subjects, because of inadequate technologies and limited knowledge of the cognitive mechanisms underlying such tasks.

Lederman and Klatzky [7] performed two experiments trying to understand the link between hand movements and object exploration with haptic feedback in blindfolded subjects. Their results show that in free exploration the procedure used to explore objects is necessary but not sufficient, i.e. haptic exploration is efficient in the exploration of 3D objects but needs to be further developed in order to perceive spatial layout and structure, e.g. a raised two-dimensional environments [8]. In this direction, Yu et al. [9] stated that spatial perception and proprioception are helpful cues in non visual exploration.

2.1. Orientation & Mobility aids

Many tactile-to-vision substitution systems have been developed and discussed in the literature [10], thanks to their low invasiveness and low costs. Several technological solutions have been proposed over time, each affected by some side effects, such as a reduced touchable surface

as in the Optical to TActile CONverter (Optacon) device (see [11] for an historical review), and disruptive interferences between sensory modalities, mostly at the expense of audition [12]. Nevertheless, multimodal virtual reality systems have gained in computational power and reliability during the last two decades.

With regard to orientation and mobility (O&M) aids in particular, several systems have been proposed up to today. The *HOMERE* system for virtual map exploration [13] replicates a virtual white cane useful in exploration tasks of predefined paths. Unfortunately, it requires a large and expensive installation and a cumbersome audio surround system. Lahav et al. [1] performed some preliminary and qualitative studies on map exploration by blind subjects. Their aim was to extract main requirements for the development of haptic virtual environments. Results showed a reduction in the exploration time for participants who mastered navigation in unknown virtual spaces compared to a control group who explored a real space only.

A more complete and complex desktop virtual reality environment is *BlindAid* [14] based on a Phantom Desktop¹ device for haptic rendering. Actions and commands are borrowed from standard human-computer interaction practice, such as zooming, scrolling, pausing, undoing, etc. *BlindAid* indeed tries to accelerate user's exploration in a spatial learning process. Unfortunately, there is no evidence nor user-centered evaluation about commands and virtual haptic textures; moreover, spatial audio was rendered through headphones with generic Head-Related Transfer Functions (HRTFs) of a KEMAR mannequin without a customization procedure. These limitations lead to a long training process.

2.2. The TActile Mouser

In the TAMO device, information about local heights of virtual objects is provided by a lever, placed at the approximate location where a mouse-wheel is commonly found. Users keep a fingertip in contact with the lever and the stepper-motor raises it proportionally to a virtual height to be rendered (an example is depicted in Fig. 1). In practice, the haptic navigation metaphor is similar to actively exploring objects with only one fingertip and kinesthetic exploration cues. When the pointer reaches

¹A 6-DOF position input/3-DOF force output device with a stylus grip. Phantom's website: <http://www.sensable.com>

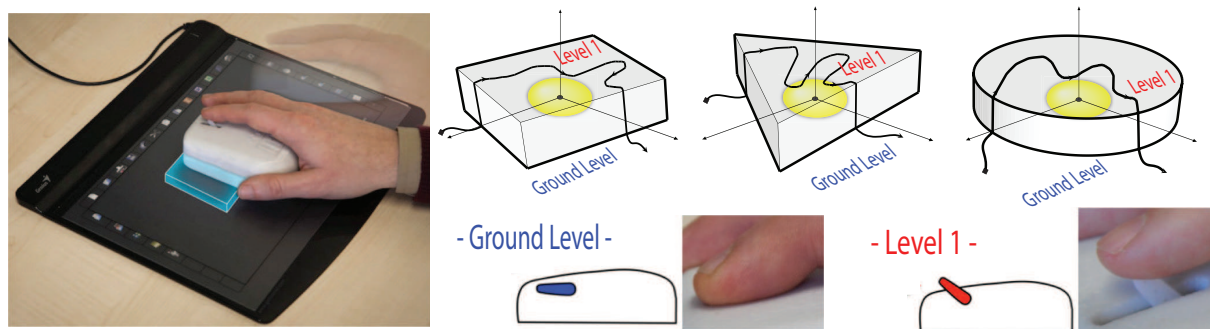


Fig. 1: The TAMO device within a single frame of exploration for a virtual object, depicted in light blue (on the left). Exploration trajectories (solid black lines) of three virtual objects with anchor sounds sketched in yellow (top). Two states of the TAMO's lever according to time-varying positions on the map (bottom): when the finger is on top of the virtual object (Level 1) the lever goes up; when the finger does not touch the virtual object (Ground Level), the lever goes down.

virtual objects on a map, the stepper signals a virtual edge of a given height. TAMO generates a *taxel*² for each pixel belonging to the working area, similarly to a tactile bas-relief representation.

The following motivations make this device suitable for our purposes:

- the combination of haptic feedback and active exploration is able to guide recognition of simple geometries;
- minimalist feedback guarantees a rapid learning curve and has relevant practical implications for end-users;
- from a commercial point of view, the TAMO provides a portable and low-budget solution compared to most haptic devices (e.g. the Phantom).

However, the minimalist haptic information provided by the TAMO gives precise information on a limited area only. In past works [6], it has been shown that user strategies are naturally deployed to integrate, with hand motion, sequences of local haptic information in a global mental map. Alternatively, global information may be delivered through complementary sensory channels. Since the auditory system seems to be the vicarious sense used by blind subjects to acquire global information about the surrounding environment (a striking example in this respect is human echolocation [15]), in this paper we explore the use of audio to deliver spatial information.

²A single tactile atomic information unit.

3. THE EXPERIMENT: OBJECT RECOGNITION

Recent investigations [6, 16] discussed on which combination of factors among cognitive load, information acquisition rate and random/identifiable strategy, affects the quality of map construction by means of local tactile information. Following a similar approach, this experiment investigates exploration strategies in a recognition task using multimodal stimuli.

Subjects were asked to explore a virtual map using the TAMO. They had to recognize simple virtual objects with basic geometric shapes as quickly as possible. Objects were placed in the center of the virtual working area, i.e. the space limited by the tablet's edges in Fig. 1. Depending on the feedback condition, a spatialized anchor sound was synthesized in the center of the tablet and served as a global orientation cue. A specially designed spatial audio engine (described next) renders the position of the anchor sound relative to the listener. An egocentric view of the virtual map was provided in which the pointer corresponds to listener's head. For example, moving the TAMO from the center to the right side of the tablet would cause the acoustic anchor sound to be spatialized towards the left, as if the head of the listener were virtually placed on the finger, like in *ears in hand* metaphor proposed by Magnusson et al. [17]. The anchor sound was rendered in two dimensions (i.e. azimuth and distance). In particular, the rendering was not influenced by local haptic height information.

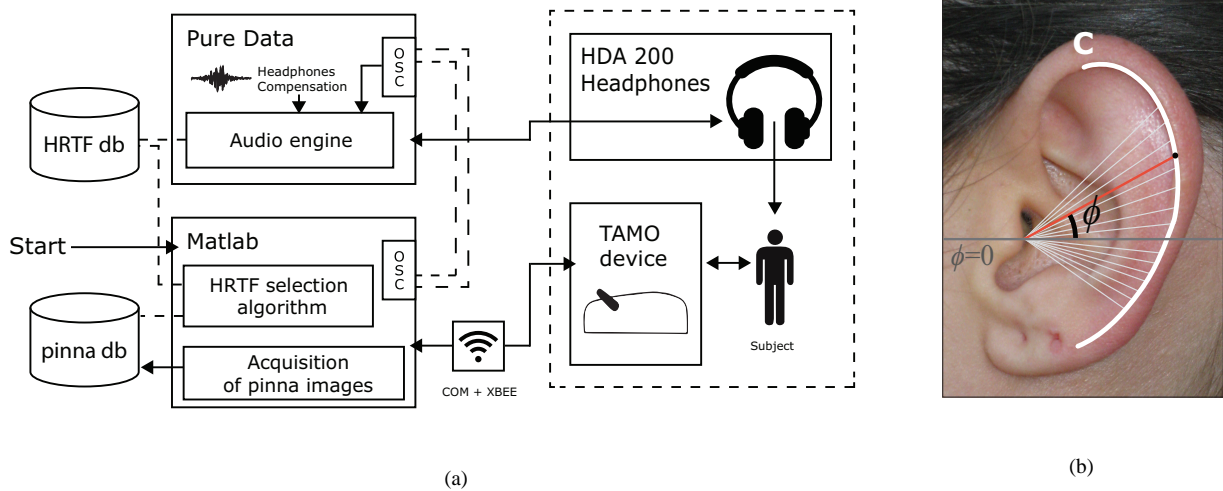


Fig. 2: (a) Simplified scheme of the experimental apparatus. (b) Pinna contour of a subject.

3.1. Participants and apparatus

Nine subjects (6 males and 3 females) whose age varied from 21 to 40 (mean 29, SD 5.7), took part to our preliminary experiment. All subjects reported normal hearing according to the adaptive maximum likelihood procedure [18]. They had different levels of expertise in psychophysical experiments: only two subjects had previous experience with multimodal experiments, all the others were naïve subjects.

The experiment was performed in a silent booth. Sennheiser HDA 200 headphones³ were plugged to a Roland Edirol AudioCapture UA-101 external audio card working at 44.1 kHz sampling rate. The TAMO device and its components (hardware and communication modules) were integrated in a new prototyped interface for haptic rendering. See Fig. 2(a) for a schematic view.

3.2. Multimodal stimuli

Depending on the stimulus presentation, a virtual sound source was placed at the center of the map (see Fig. 1). The spatial sound was rendered through headphones according to the relative position between the TAMO pointer and the anchor sound. Each auditory stimulus was a continuous train of repeated 40-ms Gaussian noise

³These dynamic closed circumaural headphones offer an effective passive ambient noise attenuation, and are able to mask mechanical sounds from TAMO.

bursts with 30 ms of silence between each burst. A similar stimulus was employed in localization tasks [19] and in navigation tasks [2]. The maximum measured amplitude of a raw stimulus at the entrance of the ear canal was set to 60 dB(A) and subjects could adjust this default level in order to obtain a subjectively comfortable level. When the device exceeded the working area, auditory feedback was stopped.

For each subject, an image-guided selection procedure (described in [20]) was performed in order to choose a HRTF set that optimizes spatial impression. An image of his/her pinna was used to compute a mismatch function between manually traced pinna contours and notch central frequencies of 45 individual HRTF sets in the CIPIC database [21]. The HRTF set providing minimum mismatch was selected for the audio rendering. The following equation formalizes the above mentioned metrics, which is derived from the *pinna reflection model* discussed in [20]:

$$m = \frac{1}{|\phi|} \sum_{\phi} D_c(\phi) \frac{|D_c(\phi) - d_c(\phi)|}{D_c(\phi)d_c(\phi)}, \quad (1)$$

where ϕ denotes the elevation angle (spanning from -45° to 45° with $|\phi|$ the number of available measurements in this range), $d_c(\phi)$ denotes the actual distance between the reflection point on the pinna contour C and the entrance of the ear canal (estimated from the pinna image), and $D_c(\phi)$ denotes the theoretical distance that

would generate the first spectral notch in the considered HRTF set. See Fig. 2(b) for an example.

Distance was rendered through an inverse square law on sound attenuation level. The sound level decayed from a maximum sound pressure level when the pointer covered approximately the sound source position, to a minimum audible sound level set to the farthest reachable position along the tablet borders. A 25 px-radius (≈ 12 mm) circle neighborhood around the anchor sound was defined in which the auditory feedback remained constant (i.e., frontal direction with azimuth $\theta = 0$, and maximum level of sound intensity).

Auditory stimuli were filtered through the selected HRTF set, and a headphone compensation filter obtained with the algorithm presented by [22], applied to headphone responses measured on a KEMAR mannequin without pinnae. Although headphone compensation was not individual, such kind of processing guaranteed effective equalization of the headphones up to 8 – 10 kHz on average and it simulated a realistic application scenario where it is not feasible to design personal compensation filters.

Finally, the proposed multimodal virtual environment is able to haptically render the presence of virtual objects on a 210×297 -mm sensing tablet. TAMO's lever moves from ground horizontal position $\varphi_0 \approx 0^\circ$, to a nearly vertical position, corresponding to a rotation of $\varphi_{max} \approx 80^\circ$ (additional information about TAMO is available in [6]). All maps were surrounded by virtual walls rendered with φ_{max} . When the device moved outside the working area the lever moved alternatively from φ_{max} to $\varphi_{max} - 26^\circ$ at a fixed refresh rate in order to signal that subjects crossed the boundaries delimited by virtual walls.

3.2.1. Multimodal integration

Since our system exploits different software/hardware components (e.g. COM ports, X-Bee[®] adapters, mouse device, HRTF interpolation, etc.), it is crucial to verify if real-time constraints are satisfied, with particular attention to multimodal stimulus synchronization within a coherent perceptual time window of integration. Since signals coming from different sensory modalities have different time-of-arrivals and processing time in the brain, a temporal window of about 200 ms ensures multisensory integration and enhancement [23]. Therefore, the two modalities produce a unitary percept.

In order to choose the refresh rate for the rendering process, the latency of audio and tactile stimuli was measured by placing two condenser microphones connected to a Tascam 680 at 192 kHz sampling rate, one at the headphones coupler and the other near the TAMO's lever. A single *taxel-plus-spatialized noise burst* was rendered and an average 68 ms delay (< 200 ms) was measured, thus guiding the choice of a 80 ms refresh rate for a consistent synchronization of the multimodal rendering loop.

3.3. Procedure

A brief tutorial session introduced the experiment. Subjects were verbally informed that they had to identify an unknown object in a virtual map using audio (headphones) and haptics (a tactile mouse). At the same time, the exploration metaphor of the TAMO device was described. During the exploration task, subjects were blindfolded and had to keep a fixed head pose. The latter indication helped subjects to keep a coherent auditory rendering. The experimenter guided blindfolded subjects to the silent booth and subsequently led them towards to a starting position of each trial, i.e. TAMO device in the middle of the bottom edge of the working area.

Each trial was completed when subjects verbally identified the object or after a maximum allowed amount of time (set to 150 s); they could guess, but no suggestion about performance was provided until the right answer was given. If subjects were not able to identify the object, the trial was concluded with a negative outcome.

Basic virtual objects, unknown to the subjects, were: a parallelepiped with triangular base, a parallelepiped with square base, and a cylinder (see Fig. 1). Object sizes were set according to a previous study [6] and they were all of equal virtual height ($\varphi_0 + 18^\circ$), yielding TAMO movements that span from ≈ 100 mm to 150 mm on virtual objects. With these constraints, object shapes had an area larger than 50% of the workspace.

Each object was presented in two conditions: (i) TAMO, unimodal haptic condition, and (ii) TAMO + 2D audio, bimodal condition. Therefore, a total of six stimuli (3 objects \times 2 conditions) were considered. Presentation sequences were arranged in latin square order with respect to object shape. Feedback conditions were presented alternatively in order to minimize learning effects.

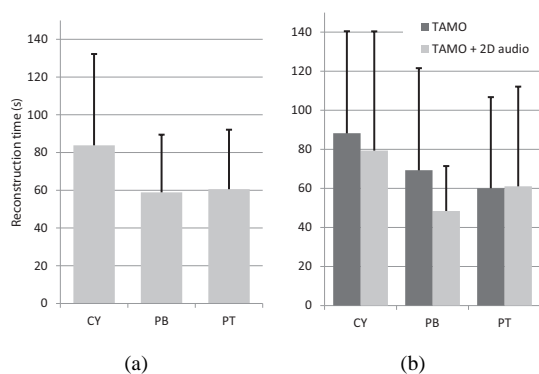


Fig. 3: Mean and standard deviation of recognition time (a) per virtual object and (b) refined per stimulus condition across 9 subjects. Labels to identify virtual objects: CY - cylinder, PB - parallelepiped with square base, and PT - parallelepiped with triangular base.

4. RESULTS & DISCUSSION

A preliminary analysis revealed an average recognition time of 72,25 s in the TAMO condition and of 62,87 s in the bimodal condition, exhibiting a slight difference denoted by standard deviations of 29,8 s and 35,9 s, respectively. Moreover, failed recognitions were sporadic.

Generally speaking, parallelepipeds were the easiest objects to be recognized, while the cylinder was the most difficult one (see Fig. 3(a)). Furthermore, the spatialized anchor sound had a limited effect on average recognition time except when subjects explored a parallelepiped with square base resulting in an increase of 30% in performance between TAMO and TAMO + 2D Audio. Moreover, the two conditions exhibit standard deviations of 46,62 s and 23 s, respectively.

In addition to the above quantitative analysis of recognition times, a qualitative analysis of subjects' exploration strategies can reveal differences between unimodal and bimodal conditions. The following discussion is focused on the exploration of parallelepipeds with square base because more differences are detectable. First of all, the same two main strategies reported by [6] were found:

1. Grid-scan: exploring the map on a grid and moving along vertical and horizontal lines (Fig. 4(c));
2. Z-scan: following object contours trying to cross the edges orthogonally or diagonally (Fig. 4(e)).

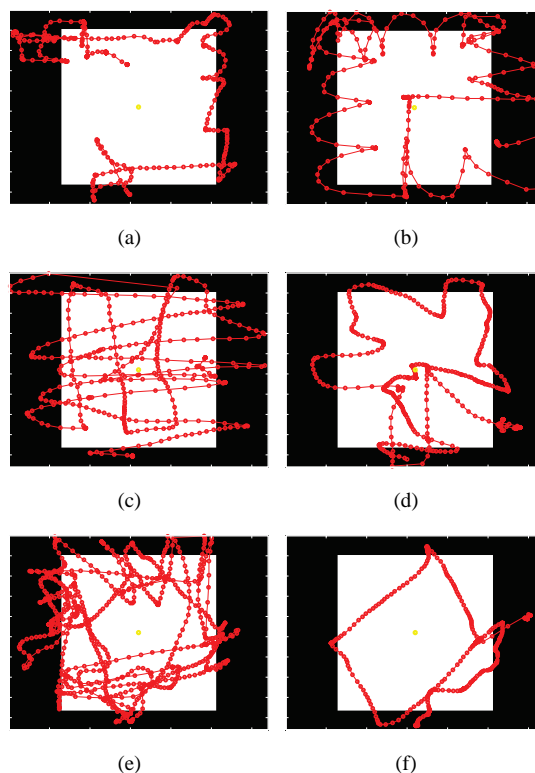


Fig. 4: Exploration of a parallelepiped with square base by three subjects (rows), in unimodal and bimodal conditions (columns). Trajectories are shown in red.

Surprisingly, a third strategy could be identified here only when the acoustic anchor was displayed: subjects followed object contours while maintaining a controlled distance from the center of the object, which corresponded to the location of the anchor sound. This behavior is depicted in Fig. 4 (second column) for three subjects. We suggest that this new strategy reduces the amount of time spent outside the object, guiding subjects through an optimized exploration. However, it is still interesting to analyze micromotion (small precise movements) and macromotion (macroscopic quick changes on the map) of subject movements. Further analysis could correlate those movement features to changes in feedback and strategies.

Such analysis could also provide additional insight about the use of anchor sounds in exploration tasks that combine far- and near-field activities. We speculate that the exploration of large environments may be supported

more effectively by using a fixed beacon sound. On the other hand, if the working area is smaller, users may benefit from directly controlling the positions of the anchor sound and using auditory spatial cues in order to create several views of the object that is being explored. In addition, a more detailed analysis of the exploration paths can reveal criticalities in HRTF representation, object complexity and anchor locations.

Possible intrinsic limitations of the haptic rendering should also be analyzed more carefully. The poor performance exhibited by all subjects in recognizing the cylinder may be due to insufficient training, but may also hint to the fact that the TAMO is only able to render simple shapes. In this respect, tactile object properties could be explored (e.g. roughness, convexity), in addition to properties that are mostly related to proprioception (such as shape).

Finally, changing the acoustic properties of the anchor may influence the cognitive load, therefore performances. Specifically, in this study we did not consider reflective sounds on virtual walls. Early reflections from the environment are known to substantially contribute to distance and environment perception; a deeper study that relates haptic shapes and acoustic features is necessary in order to improve performances of exploration/recognition task in virtual reality.

5. CONCLUSIONS

The multimodal system proposed in this paper provides an experimental testbed for the use of spatial audio to convey global environmental information for navigating in virtual spaces. Since the local haptic information is similar to tactile maps widely used in rehabilitation and O&M learning activities, an object recognition task with blindfolded subjects provided preliminary results on how anchor sounds can improve performances and modifications in subjects' exploration strategies.

The proposed exploration metaphors using haptics and audition mimic strategies that are naturally adopted in the real world and therefore provide an ecological approach to sensory substitution of vision for the considered exploration and recognition task.

In order to establish reliable design guidelines for complex O&M environments that make use of spatial audio, and to investigate higher-level cognitive functions,

such as spatial cognition, it is important to define novel efficiency metrics that take into account subjects' performance and exploration strategies. Such metrics are needed to guide the life-cycle of user-centered applications for visually-impaired people. The results presented in this work provide preliminary indications towards the definition of such metrics.

6. ACKNOWLEDGMENT

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

7. REFERENCES

- [1] Orly Lahav and David Mioduser. Exploration of unknown spaces by people who are blind using a multi-sensory virtual environment. *Journal of Special Education Technology*, 19(3):15–23, 2004.
- [2] Bruce N. Walker and Jeffrey Lindsay. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2):265–278, 2006.
- [3] Amandine Afonso, Alan Blum, Brian Katz, Philippe Tarroux, Grégoire Borst, and Michel Denis. 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*, 38(5):591–604, 2010.
- [4] Lauri Savioja, Jyri Huopaniemi, Tapio Lokki, and Ritta Väänänen. Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, 47(9):675–705, 1999.
- [5] G. C. Vanderheiden. Nonvisual alternative display techniques for output from graphics-based computers. *Journal of Visual Impairment and Blindness*, 83(8):383–90, January 1989.
- [6] Luca Brayda, Claudio Campus, Ryad Chellali, Guido Rodriguez, and Cristina Martinoli. An investigation of search behaviour in a tactile exploration task for sighted and non-sighted adults. In *Proceedings of the 2011 annual conference extended*

- abstracts on Human factors in computing systems*, CHI EA '11, page 2317–2322, Vancouver, BC, Canada, 2011. ACM. ISBN 978-1-4503-0268-5.
- [7] Susan J. Lederman and Roberta L. Klatzky. Hand movements: A window into haptic object recognition. *Cognitive Psychology*, 19(3):342–368, 1987.
- [8] Lochlan E. Magee and John M. Kennedy. Exploring pictures tactually. *Nature*, 283(5744):287–288, January 1980.
- [9] Wai Yu and Stephen Brewster. Comparing two haptic interfaces for multimodal graph rendering. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proceedings. 10th Symposium on*, page 3–9, 2002.
- [10] Kurt A. Kaczmarek, John G. Webster, Paul Bach-y-Rita, and Willis J. Tompkins. Electrotactile and vibrotactile displays for sensory substitution systems. *Biomedical Engineering, IEEE Transactions on*, 38(1):1–16, 1991.
- [11] William Schiff and Emerson Foulke. *Tactual Perception: A Sourcebook*. Cambridge University Press, March 1982. ISBN 9780521240956.
- [12] René Farcy, Roger Leroux, Alain Jucha, Roland Damaschini, Colette Grégoire, and Aziz Zogaghi. Electronic travel aids and electronic orientation aids for blind people: technical, rehabilitation and everyday life points of view. In *Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Technology for Inclusion*, page 12, 2006.
- [13] A. Lecuyer, P. Mobuchon, C. Megard, J. Perret, C. Andriot, and J.-P. Colinot. HOMERE: a multimodal system for visually impaired people to explore virtual environments. In *Virtual Reality, 2003. Proceedings. IEEE*, pages 251–258, March 2003.
- [14] D.W. Schloerb, O. Lahav, J.G. Desloge, and M.A. Srinivasan. BlindAid: virtual environment system for self-reliant trip planning and orientation and mobility training. In *Haptics Symposium, 2010 IEEE*, pages 363–370, March 2010.
- [15] Lore Thaler, Stephen R. Arnott, and Melvyn A. Goodale. Neural correlates of natural human echolocation in early and late blind echolocation experts. *PLoS ONE*, 6(5):e20162, May 2011.
- [16] L. Brayda, C. Campus, and M. Gori. Predicting successful tactile mapping of virtual objects. *IEEE Transactions on Haptics*, 6(4):473–483, October 2013.
- [17] Charlotte Magnusson, Henrik Danielsson, and Kirsten Rasmus-Gröhn. Non visual haptic audio tools for virtual environments. In David McGookin and Stephen Brewster, editors, *Haptic and Audio Interaction Design*, number 4129 in Lecture Notes in Computer Science, pages 111–120. Springer Berlin Heidelberg, January 2006. ISBN 978-3-540-37595-1, 978-3-540-37596-8.
- [18] D. M. Green. A maximum-likelihood method for estimating thresholds in a yes-no task. *J. Acoust. Soc. Am.*, 93(4):2096–2105, April 1993.
- [19] Brian F. G. Katz, Slim Kammoun, Gaëtan Parsehian, Olivier Gutierrez, Adrien Brilhault, Malika Auvray, Philippe Truillet, Michel Denis, Simon Thorpe, and Christophe Jouffrais. NAVIG: augmented reality guidance system for the visually impaired: Combining object localization, GNSS, and spatial audio. *Virtual Reality*, 16(4):253–269, June 2012.
- [20] Michele Geronazzo, Simone Spagnol, Alberto Bedin, and Federico Avanzini. Enhancing vertical localization with image-guided selection of non-individual head-related transfer functions. In *In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, pages 4496–4500, Florence, Italy, May 2014.
- [21] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The CIPIC HRTF database. In *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, page 1–4, New Paltz, New York, USA, October 2001.
- [22] Alexander Lindau and Fabian Brinkmann. Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings. *J. Audio Eng. Soc.*, 60(1/2):54–62, 2012.
- [23] Nicholas P. Holmes and Charles Spence. Multi-sensory integration: Space, time, & superadditivity. *Curr Biol*, 15(18):R762–R764, September 2005.