Wayfinding with Simulated Prosthetic Vision: Performance comparison with regular and structure-enhanced renderings

Victor Vergnieux, Marc Macé, Christophe Jouffrais

HAL Id: hal-01150335
https://hal.archives-ouvertes.fr/hal-01150335

Wayfinding with Simulated Prosthetic Vision: Performance comparison with regular and structure-enhanced renderings

Victor Vergnieux, Marc J.-M. Macé and Christophe Jouffrais

Abstract — In this study, we used a simulation of upcoming low-resolution visual neuroprostheses to evaluate the benefit of embedded computer vision techniques in a wayfinding task. We showed that augmenting the classical phosphene rendering with the basic structure of the environment - displaying the ground plane with a different level of brightness - increased both wayfinding performance and cognitive mapping. In spite of the low resolution of current and upcoming visual implants, the improvement of these cognitive functions may already be possible with embedded artificial vision algorithms.

I. INTRODUCTION

A. Visual neuroprosthesis

Visual neuroprosthesis and in particular retinal implants have been developed within the last decades to restore partial visual perception in blind people. In a retinal implant, an array of microelectrodes is attached to the retina. Ideally, each microelectrode can electrically elicit one visual percept in the form of a white or yellow dot called a phosphene. A maximum of 3 to 10 levels of brightness can be perceived when the intensity of the stimulation is modulated [1]. Since early in the development of visual prosthesis, several groups have been investigating the potential benefits from such implant via simulation. Even though Simulated Prosthetic Vision (SPV) systems cannot replace experiments with implantees, they are used to assess the functional capacities that could be restored with future implants (see e.g. [2]).

B. Navigation & Prosthetic vision

SPV studies have mostly focused on reading capabilities or object recognition while orientation or navigation have been less studied. The first study on this topic [3] used a narrow field of view (1.7°) and a very high resolution (625 phosphenes) that was relatively unrealistic but very stimulating for further studies. Following SPV studies have explored some aspects of mobility such as obstacle avoidance [4]–[7] and to our knowledge, only one study has investigated the restoration of higher level functions such as wayfinding [8]. The main results of those studies on mobility were that the minimal number of phosphenes required to navigate with ease in an unknown environment is a few hundreds.

As a consequence, retinal implants recently introduced on the market, such as the Argus II (60 electrodes) from Second Sight Medical Products (Sylmar, California, USA), do not have a sufficient resolution to actively navigate in an unknown environment, and hence promote cognitive mapping. The more obvious solution would be to increase the array resolution but several technical problems are slowing down this evolution [9]. Waiting for higher resolution implants, another solution is to enhance the perception provided through low resolution implants with pertinent navigation cues selected in the environment. Most of the studies cited above aimed to enhance the contrast between phosphenes to highlight the structure of the environment or some obstacles located in front of the user. For instance, methods have been applied to automatically extract the ground plane in the camera image and highlight it in order to enhance contrast with the walls [10]. Similar methods were applied to all surface boundaries with more efficient algorithms [11]. Although the feasibility of real-time structure enhancement has been demonstrated, behavioral and cognitive benefits to the user have not been clearly addressed. In this study, we used SPV to realize a navigation task in an unknown virtual environment. We specifically assessed the improvement of wayfinding in relation to contrast enhancement of the visual scene. We also evaluated the quality of the maps that the subjects drew after navigation with standard vs. augmented environment renderings. The behavioral task was a game inspired by [12], which encourages subjects to navigate and memorize a mental map of the environment.

This study had two objectives: evaluate the wayfinding performance, and assess the quality of the cognitive mapping built during navigation in an unknown environment, both with a simulated low resolution visual implant.

II. METHODS

A. Subjects

12 subjects, (6 males and 6 females aged 23 to 39), participated in the experiment. Every participant had normal or corrected to normal vision. This experiment was conducted according to the ethical recommendations of the Declaration of Helsinki and was approved by a local ethical

V. Vergnieux is with Université de Toulouse ; UPS ; IRIT 118 route de Narbonne, F-31062 Toulouse Cedex 9, France. M. J.-M. Macé and C. Jouffrais are with CNRS ; IRIT F-31062 Toulouse, France. (Corresponding author: Victor Vergnieux; phone: +33 561-556-305; e-mail: elipse_SPV@irit.fr).

committee (CLERIT) at the University of Toulouse. All subjects gave written informed consent to participate.

B. Protocol

The virtual environments were simple urban environments with medium size walls. We designed three environments containing seven segments defining three central blocks (see Fig. 1a). All the turns (3 to 5) and intersections (6) were 90° angles. In each environment, the corridors had a total length of 200 m, and the subject moved at the speed of 3 m/s. In order to equate the complexity among environments, the structure of the first one was scrambled to generate the second and third ones.

The subjects had to explore an unknown virtual maze to find and bring jewels back to the starting point (called "base"). They had 5 minutes to collect 9 jewels in each environment. The subjects had the instruction to collect the jewels as fast as possible, and to memorize the environment in order to draw the corresponding map at the end of the game. To collect a jewel, the subject had to walk through it (the jewel was then picked up) and bring it back to the base (the jewel was then secured). In order to encourage subjects to memorize the spatial configuration of the environment, only 3 jewels could be carried at the same time. Hence they had to make at least 3 journeys in order to secure all 9 jewels.
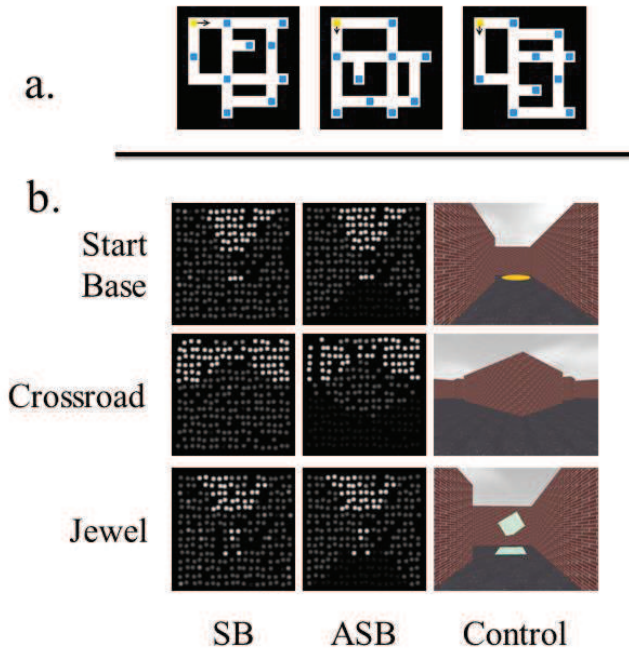

Figure 1. a. Environment maps: start base and jewels are indicated in yellow and blue respectively. The arrow represents the subject's orientation at the beginning of the task. b. Samples of the three rendering conditions used in this study.

Three conditions were compared: two SPV conditions and one control condition (Fig. 1b). In the control condition, the subject saw the environment as originally displayed by the 3D engine. In the standard SPV rendering -called scoreboard (SB)-, the resolution of the camera image was reduced to 15*18 pixels and 4 levels of gray to match the resolution and color depth of the simulated implant. In the second SPV rendering -called augmented scoreboard rendering (ASB)-, the rendering was similar to SB but the ground was darkened to increase its contrast with the walls.

Before playing the game, the subjects were trained to use each rendering by collecting 3 jewels in a very simple learning environment. They started the experiment when they declared being ready to do so. For each rendering condition (SB, ASB, Control), subjects played the game 3 times, in the same environment. The order of the three conditions and the condition/environment associations were counterbalanced across subjects. After completing each condition (block of three games), they had to draw the corresponding map, including the configuration of the corridors, and the location of the base and jewels. Subjects were previously informed that they will be asked to draw the map at the end of the game.

After completing the whole protocol, participants were invited to fill in a form with 5 main questions (on 7-points Likert scales). In this form, they had to evaluate the pleasure and difficulty to play the game with each rendering. They used the same scale to mention if the renderings were usable and provided enough information to complete the task. They finally quoted their own frequency of videogame playing.

C. Material

The virtual environment (VE) was displayed via a head mounted display (HMD) (NVisor SX-60 HMD, NVIS Inc., Reston, USA) with a resolution of 1280*1024 pixels and 44*34 degrees of visual angle. The orientation of the virtual camera was attached to the orientation of the head with a motion capture system (OptiTrack, Natural Point, USA). As a result, subjects could look around in the VE when moving the head. As subjects were seating, they could not rotate their body and rotation was provided in the VE using the left and right arrow keys of a keyboard. The virtual environments were generated and animated using the open source Irrlicht 3D Engine (v1.8). A basic audio feedback was provided with step sounds and a wall bump sound.

The different renderings were displayed in the HMD with a home-made SPV system. The simulated array had a resolution of 15*18 electrodes, which corresponds to the intended resolution of the next generation of implants (e.g. Second Sight Argus III). The phosphene array covered 9*9 degrees of visual angle. A typical dropout rate of 10% was used to simulate the electrodes that did not elicit any phosphene. The phosphene appearance was based on observation during clinical trials [1]. They were circular dots with a Gaussian luminance profile displayed on a greyscale with 4 levels. Further details about the system are available in [13].

D. Data and statistical analysis

As described before, the aim of the game was to secure 9 jewels within 5 minutes. If the 9 jewels were retrieved and secured before the end of the timer, elapsed time was recorded. If less than 9 jewels were secured at the end of the

5 minutes, the number of jewels was recorded. We then computed a performance index which was the average of 2 percentages:

• Percentage of jewels secured.

• Percentage of remaining time: percentage of 100% was granted for the minimum time needed to capture the 9 jewels (computed for each environment); and percentage of 0% was granted if the timer ended before bringing back the 9 jewels.

Hence, a performance index of 50% meant that the subject had collected the 9 jewels in exactly 5 minutes. A lower performance index meant a failure in securing all the jewels in 5 minutes. A performance index higher than 50% meant that the subject succeeded in bringing back all the jewels, and in less than 5 minutes.

We used the R software (v3.0.1) for statistics. Because the data did not follow a normal distribution, we used Wilcoxon tests for all comparisons (paired data). Statistical significance was set to $p < 0.05$, and adjusted with a Bonferroni correction when needed.

III. RESULTS

A. Quantitative results

With the SB rendering, the number of secured jewels was only 5.2 (SD=2.3) in average, and hence significantly inferior (Z=4.4, $p < 0.001$) to the average number of secured jewels with the ASB rendering (7.3 ±2.3) (see Fig. 2). Likewise, average game duration was significantly higher (Z=3.8, $p < 0.001$) with the SB rendering (295 ±21s) than with the ASB rendering (265 ±47s). Accordingly, the performance index (PI) significantly differs (Z=4.6, $p < 0.001$) between the SB rendering (30.5% ±16.7) and the ASB rendering (51.2% ±24.2). The averaged PI in the control condition (81.5% ±16.5) was significantly higher than the PI in both SPV conditions (SB: Z=5.2, $p < 0.001$; and ASB: Z=5.0, $p < 0.001$).
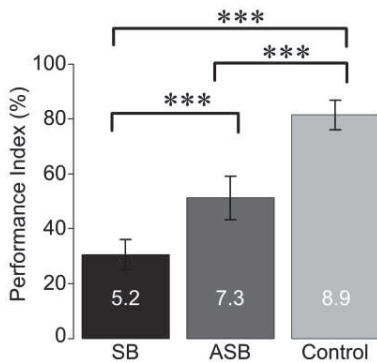
Figure 2.   Performance index per condition: Scoreboard (SB), Augmented Scoreboard (ASB), and Control condition. Mean number of secured jewels in each condition is written in white within the columns. Error bars are IC95.

Three judges were asked to evaluate the quality of the maps of the environment that the subjects drew after each condition. The judges were not involved in the experiment, and were not informed about the objectives. They were instructed to evaluate the similarity between each drawing and the real map of the environment without considering distance distortion. They could mark the drawings between 0 (drawing and map completely different) and 10 (drawing and map perfectly consistent). Examples of drawings are shown in Fig. 3.

The mean drawing score for the SB condition was 2.7 (SD=2.3) and was significantly lower than the mean drawing score for the ASB condition (4.9 ±3.1; Z=3.1, $p < 0.01$). The mean drawing score for the control condition (7.8 ±1.7) was significantly higher than the score of both SPV conditions (SB: Z=5.7, $p < 0.001$, and ASB: Z=4.5, $p < 0.001$). No statistical difference was found between the three environment configurations shown in Fig. 1a.
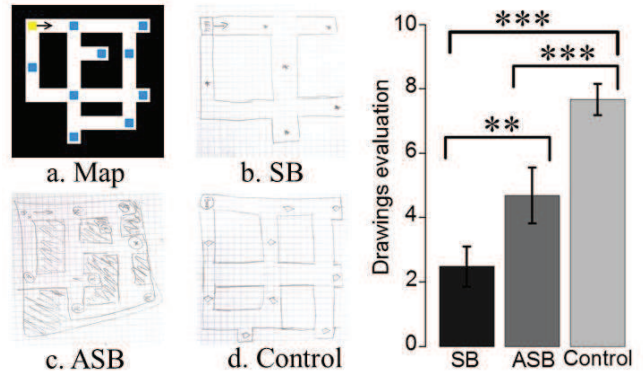
Figure 3.   Left: Map and drawing examples (from different subjects) a. Map of environment #1; b. Drawing after SB condition; c. after ASB condition; d. after Control condition. Right: Mean drawing scores per condition (n=12 subjects). Error bars are IC95.

B. Qualitative Results

1) Questionnaire

11 out of 12 subjects evaluated the task easier and more pleasant with the ASB rendering than with the SB rendering. The same proportion of subjects granted the ASB rendering to be more usable than the SB rendering. 10 subjects reported that the ASB rendering was providing enough information to complete the task. Only 3 subjects had the same opinion about the SB rendering.

The 2 subjects with the best ASB performance index were the only ones who consider themselves as "passionate gamer". The 4 subjects who did not play video games at all were the 4 subjects with the worst performance indexes with the ASB and the SB rendering.

IV. DISCUSSION

In this experiment, we assessed the ability to navigate and build a mental representation of the travelled environment with two different simulated phosphene renderings. These renderings correspond to low resolution implants that will soon be evaluated within clinical trials.

We first observed that the control rendering provided the subjects with enough landmarks and spatial cues to complete the game and draw a consistent map of the environment. We

then observed that the subjects were not able to complete the game with the "standard" (SB) rendering. Despite a strong contrast between the sky and the walls (see Fig. 1), they were unable to bring all the jewels back to the base, nor draw a map of the explored environment. In comparison, the ASB rendering brought very crude additional cues concerning the path (only the floor plane was rendered differently), but these cues were sufficient to complete both tasks more quickly and accurately. In average, subjects brought back 40% more jewels in ASB condition than in SB condition and their drawings of the environment received scores 80% higher. These large differences between the two conditions may be explained by a combination of factors. The first explanation is related to spatial updating which is necessary to understand one's own displacements. When subjects were virtually walking in a corridor, or turning around, the enhanced floor cues were probably useful to better estimate self-motions. A second explanation could be related to the better detection and understanding of intersections and turns in the ASB condition, which could then be used as reliable landmarks to navigate towards the different parts of the maze but also to build a reliable mental representation of the environment [14].

The task used in this experiment is relatively different from the mobility task used in previous SPV studies. Most of them were performed in indoor environments and focused on detecting doors or obstacles in tasks that are closer to mobility than wayfinding (e.g. [4]). Here, the cognitive functions that were improved during wayfinding in a virtual outdoor environment were related to path integration, self-orientation and cognitive mapping. This would be a realistic solution to help people with visual implants navigate actively and independently.

A limitation of this study is that the designed environments and the interaction techniques (keyboard) are very similar to those encountered in videogames, which may favored the gamers. Although it does not change the observation that augmented rendering improves cognitive functions, it may be useful to address this issue. The solution may be to adopt a whole body interaction technique in order to move in the virtual environment, or to perform this experiment in a real outdoor environment with a mobile SPV.

V. Conclusion

In this SPV experiment, we obtained two main results. First, we confirmed the plausibility to actively explore an unknown environment with the upcoming generation of low-resolution retinal implants (15*18 electrodes). However, wayfinding and cognitive mapping will probably still be problematic with standard rendering. Embedded real-time artificial vision methods that are currently available can provide crude but useful cues about the environment. We showed that such a simple and realistic cue -augmented contrasts between the floor and the walls- would improve orientation, wayfinding and cognitive mapping.

More generally, we suggest that enhancing the low resolution rendering of visual implants with specific cues such as landmarks, obstacles ([4]) or waypoints will be useful for restoring the autonomy of implanted patients. To increase further the usability of the system, we could think of a computer vision-based system that would detect other types of objects in the environment such as faces, blocks of text [15], etc., with a specific interaction technique to switch between these different detection modes.

VI. References

[1] S. C. Chen, G. J. Suaning, J. W. Morley, and N. H. Lovell, "Simulating prosthetic vision: I. Visual models of phosphenes.," *Vision Res.*, vol. 49, no. 12, pp. 1493–1506, Jun. 2009.

[2] S. C. Chen, G. J. Suaning, J. W. Morley, and N. H. Lovell, "Simulating prosthetic vision: II. Measuring functional capacity.," *Vision Res.*, vol. 49, no. 19, pp. 2329–43, 2009.

[3] K. Cha, K. W. Horch, and R. A. Normann, "Mobility performance with a pixelized vision system.," *Vision Res.*, vol. 32, no. 7, pp. 1367–72, 1992.

[4] J. A. Dowling, W. W. Boles, and A. J. Maeder, "Simulated artificial human vision: The effects of spatial resolution and frame rate on mobility.," in *Proceedings Active Media technology*, 2006, pp. 138–143.

[5] J. A. Dowling, A. Maeder, and W. Boles, "Mobility enhancement and assessment for a visual prosthesis.," in *Medical Imaging*, 2004, pp. 780–791.

[6] C. McCarthy and N. Barnes, "Time-to-contact maps for navigation with a low resolution visual prosthesis.," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE.*, 2012, pp. 2780–3.

[7] N. Barnes, P. Lieby, H. Dennet, J. Walker, C. McCarthy, N. Liu, and Y. Li, "Investigating the role of single-viewpoint depth data in visually-guided mobility," *J. Vis.*, vol. 11, no. 11, pp. 926–926, 2011.

[8] G. Dagnelie, P. Keane, V. Narla, L. Yang, J. D. Weiland, and M. S. Humayun, "Real and virtual mobility performance in simulated prosthetic vision.," *J. Neural Eng.*, vol. 4, no. 1, pp. S92–101, 2007.

[9] S. Picaud and J.-A. Sahel, "Retinal prostheses: Clinical results and future challenges.," *C. R. Biol.*, 2014.

[10] C. McCarthy, N. Barnes, and P. Lieby, "Ground surface segmentation for navigation with a low resolution visual prosthesis.," in *Engineering in Medicine and Biology Society (EMBC), 2011 Annual International Conference of the IEEE.*, 2011, pp. 4457–60.

[11] D. Feng and C. McCarthy, "Enhancing scene structure in prosthetic vision using iso-disparity contour perturbance maps.," in *Engineering in Medicine and Biology Society (EMBC), 2013 Annual International Conference of the IEEE.*, 2013, pp. 5283–6.

[12] L. B. Merabet, E. C. Connors, M. A. Halko, and J. Sánchez, "Teaching the Blind to find their way by playing video.," *PLoS One*, vol. 7, no. 9, pp. 1–6, 2012.

[13] G. Denis, C. Jouffrais, V. Vergnieux, and M. J.-M. Macé, "Human faces detection and localization with simulated prosthetic vision.," in *Proceedings of the 2013 Conference on Human Factors in Computing Systems (CHI 2013)*, 2013, pp. 61–66.

[14] A. W. Siegel and S. White, "The development of spatial representations of large-scale environments.," in *Advances in Child Development and Behavior*, vol. 10, H. Reese, Ed. New York: Academic Press, 1975, pp. 10–55.

[15] G. Denis, M. J.-M. Macé, and C. Jouffrais, "Simulated Prosthetic Vision: improving text accessibility with retinal prostheses," in *Engineering in Medicine and Biology Society (EMBC), 2014 Annual International Conference of the IEEE.*, 2014.