Does switching between different renderings allow blind people with visual neuroprostheses to better perceive the environment?

Julien Desvergnes, Axel Carlier, Wei Tsang Ooi, Vincent Charvillat, Christophe Jouffrais

Does Switching between Different Renderings Allow Blind People with Visual Neuroprostheses to Better Perceive the Environment?

Julien Desvergnes[1,3], Axel Carlier[1], Wei Tsang Ooi[2], Vincent Charvillat[1], and Christophe Jouffrais[3]

[1] IRIT-ENSEEIHT, France
[2] National University of Singapore
[3] CNRS, France

julien.desvergnes97@gmail.com

Abstract. Visual neuroprostheses are devices that restore limited visual perception for visually impaired patients. Some of these neuroprostheses, implanted in the retina or the visual cortex include an implant, a computing device and an external camera to capture the scene. An impoverished visual perception is restored when the microstimulation of the retina or the visual cortex activates white spots called phosphenes. However, the resolution of current implants (i.e. the number and spacing of the electrodes) that have passed the clinical trial phases remains low. Such low resolution, coupled with the limited number of different colors rendered by the implants, limits the information that can be transferred. The regular rendering process used with the implants, called Scoreboard, is insufficient to support complex comprehension tasks. To allow the patient to better perceive his environment, ongoing research aims at maximizing the quality and quantity of information provided by the implant. We set up a comparative study between different renderings and we showed that providing the blind with the possibility to switch between different renderings significantly increases the understanding of the environment.

Keywords: Visual neuroprostheses, Interactive rendering, Adaptive rendering, Retinal implant, Computer vision, Blind people.

1 Introduction

According to the WHO [1], 253 million people are visually impaired: 36 million of them are blind and 217 million have moderate to severe visual impairment. Visual neuroprostheses first appeared in the 1960s [2] and have emerged as a promising technique for partially restoring vision in people with visual impairment. Over the last ten years, several implants have been placed on blind people and have been in clinical trials [3]. In this study, we compared three different prosthetic rendering modes: the Scoreboard mode (Control); a combination of semantic object segmentation and scene structure detection called "Combined", similar to a recent state of the art method [6]; and a mode

called "Switch" which allows to alternate between the Combined rendering and two other renderings where only objects or only structure appear respectively. The study is based on the analysis of the needs of blind people provided by Ratelle and Couturier [4]. Our hypothesis is that the Switch rendering mode provides a better understanding of the organization of the objects as well as the structure in an external scene than the Control (Scoreboard) and Combined rendering modes. Since it is impossible to perform such tests on implanted patients, the study was performed with a prosthetic vision simulator inspired by clinical reality. Our results, obtained on 20 subjects, show the interest of the "Switch" mode in the understanding of static scenes (images).

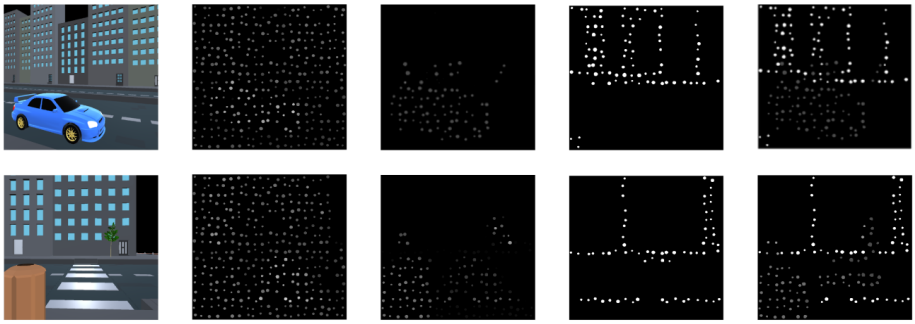

Fig. 1. Examples of prosthetic renderings: (a) shows the initial image in a virtual environment. (b) shows how (a) is rendered with Scoreboard rendering. (c) shows how (a) is rendered with a method that detects only objects, and (d) shows how (a) is rendered with a method that enhances structural information (structure enhancement). (e) is a rendering combining (c) and (d) and called Combined.

2 Related Work

2.1 Visual Neuroprostheses

Visual neuroprostheses are devices designed to restore light perception in people with partial or total blindness. They consist of a portable camera, a small computer and a matrix of electrodes that is implanted in the retina. These implants generate electrical micro-stimulations that cause the appearance of blurred points called phosphenes. Several devices have been developed and some implants have been clinically tested. Existing commercial visual neuroprosthesis systems are based on retinal implantation of a limited number of electrodes (6x10 for the Argus II developed by Second Sight [7], 21x18 for the PRIMA system developed by Pixium Vision [5]). While these neuroprostheses have improved the daily lives of many blind people [8], the restored visual perception is too weak to allow more advanced perceptual or sensorimotor processes such as navigation.

2.2 Different Phosphenic Renderings Depending on the Task

The limitations of current visual neuroprostheses pose significant problems for rendering a scene in a comprehensible manner. The historical method of "Scoreboard" rendering consists in reducing the captured image to the resolution of the implant, then converting the image to grayscale and quantifying the grayscale to the intensity level supported by the implant (Fig. 1B).

However, several studies have shown that specific image processing can improve the performance of subjects in various tasks from perception to navigation [6,9-11]. Specifically, Sanchez-Garcia et al. [6] proposed a rendering that constructs a schematic representation of indoor environments, highlighting the structural informative edges and silhouettes of segmented objects. An example of this type of rendering is proposed on Fig. 1E.

3 Prosthetic Vision Simulator

Due to the difficulty of accessing implanted patients to test different renderings, the use of a prosthetic vision simulator is common in the literature [6,9,10,12]. As its name indicates, the simulator allows to simulate different implants (size, position, resolution) in various contexts (2D or 3D visual scenes, static or dynamic).

3.1 Simulated Implant

We chose to simulate the PRIMA system developed by Pixium Vision (Paris, France), composed of a 21x18 electrodes array. Our choice is motivated by the fact that it is the most recent implant having passed satisfactory clinical trials [3]. It is also one of those that allows a visual rendering with the highest resolution to date. Indeed, the current technical limitations do not allow to have a very high number of electrodes (the Argus II has only 6x10 electrodes [13]).

A dropout rate of 10\% was applied to the electrodes to simulate non-functional or broken electrodes. A Gaussian blur was applied to the generated phosphenes, the size of the phosphenes within the same implant varied between 0.235° and 0.275° of the field of view. The spacing between two phosphenes varied between 0.55° and 0.825° of the field of view.

3.2 Phosphenic Renderings

Scoreboard rendering (Control): The image is reduced to the resolution of the implant. We then quantize the intensity into four levels of gray. Finally, we transform each rectangular area that now represents an implant electrode into a phosphene.

Combined rendering (objects + structure): First, we extract an object segmentation map from the input image, see Fig. 2-1A. As with the Scoreboard rendering, this image is scaled and quantized, see Fig. 2-2A. In parallel, we also extract the edges of the scene structure and then we scale and quantize the image, see Fig. 2-1B-2B. We then

combined 2A and 2B into a single image (Fig. 2-3). Finally, we transform each rectan gular area that now represents an implant electrode into a phosphene (Fig. 2-4).

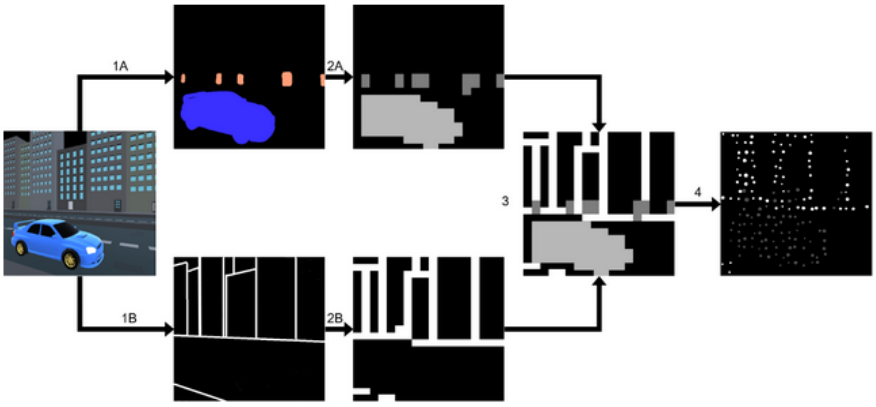

Fig. 2. Generation of the Combined rendering (objects + structure).

Switch rendering: The Switch rendering mode allows the subject to change the rendering at will. In this mode, the available renderings are the Combined rendering, the Objects rendering and the Structure rendering. The Objects rendering is obtained by transforming the image from Fig. 2-2A into a phosphenic version. The Structure rendering is obtained by transforming the image from Fig. 2-2B into a phosphenic version.

4 Material, Methods and Protocol

4.1 Hypothesis, Subjects and Experimental Conditions

We hypothesized that the ability to switch rendering types in real time to perform outdoor visual scene comprehension tasks allows subjects to perform better visually. The subjects were recruited in an engineering school and via social networks (LinkedIn, Facebook). 20 subjects, (11 men and 9 women aged from 17 to 55 years, mean: 25 years, sd: 13 years), participated in the experiment. Each participant had normal or corrected vision. All subjects gave their consent to participate in the study and allow storage of their anonymized data, in accordance with the GDPR. We used the three experimental conditions presented in section 3.2.

4.2 Protocol and Variables Analyzed

The experimental session contained three blocks corresponding to the three experimental conditions. Each block was divided into three parts including familiarization with the rendering used in the block, the test phase, and then a questionnaire regarding the rendering the subject just used in the block. During the test phase, the subject had to answer a series of 15 questions including 5 questions in three categories named "objects", "streets", "doors and crosswalks". Once an answer was selected, users could

move on to the next question. The order of the blocks and questions was randomized to limit inter-block or inter-question bias. At the end of each block, subjects were asked subjective questions about the appropriateness of the rendering used in that block. The questions focused on the pleasure experienced and the perceived usability of the rendering used in that block.

To address our hypothesis, we measured, for each subject, two variables: (i) the validity of each answer (correct / incorrect / I don't know), (ii) the response time to each question. To analyze the quantitative results we performed two-way ANOVA tests according to the model (Condition * Task * Interaction), and then used a Tukey post-hoc test to compare pairs two by two. In the figures, we used 95% confidence intervals.

5 Results

Correct answers: The two-way ANOVA showed that the number of correct answers was significantly different by rendering ($F(2,171) = 15.834$; $p < 0.0001$) and by task $F(2,171) = 81.730$; $p < 0.0001$). The interaction was also significant ($F(4,171) = 5.728$; $p < 0.001$). Switch condition is highly effective in identifying objects, but even more in understanding street patterns, see Fig. 3.

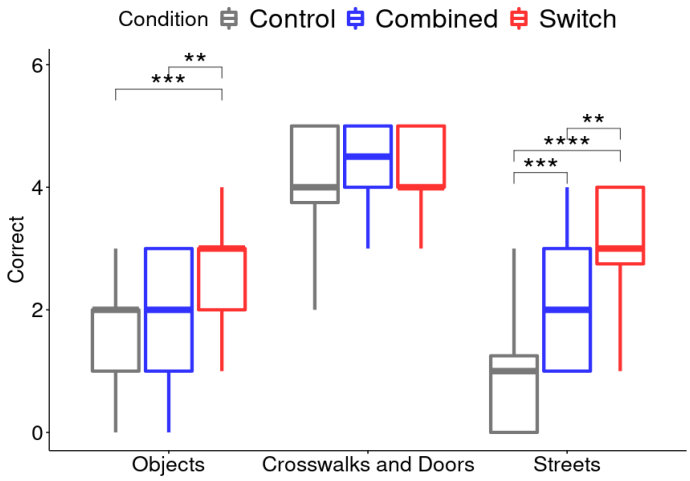

Fig. 3. Average number of correct responses per Render and per Task. Detection of crosswalks and doors is equivalent for all conditions. Identification of objects and street patterns is significantly improved with the Switch rendering. (N=20. Bars indicate 95\% confidence interval. *=0.05 ; **=0.01 ; ***=0.001 ; ****=0.00001).

"I don't know" answers: The two-way ANOVA (Condition * Task * Interaction) showed that the number of "I don't know" answers was significantly different by rendering ($F(2,171) = 10.543$; $p < 0.0001$) and by task $F(2,171) = 5.882$; $p < 0.001$). However, the interaction was not significant ($F(4,171) = 1,909$; $p = 0.111$).

Response time: The two-way ANOVA showed that the response time is significantly different according to the experimental condition ($F(2,891) = 17.245$; $p < 0.0001$) and according to the task $F(2,891) = 14.081$; $p < 0.0001$). However, the interaction was not significant ($F(4,891) = 0.465$; $p = 0.762$).

Subjective questionnaire: Participants were asked to respond with a score between 1 and 7 to a series of questions at the end of each of the three blocks. The two-factor ANOVA (Render * Task * Interaction) shows that the score was significantly different by rendition ($F(2,228) = 11.878$; $p < 0.0001$) and by task ($F(3,228) = 5.604$; $p < 0.001$). The interaction was significant ($F(6,228) = 7.537$; $p < 0.0001$). We observe that in terms of pleasure of use, the Combined and Switch renderings are significantly better rated than the Control rendering (Scoreboard). We can also observe that in terms of difficulty of use, perceived usability and the amount of information presented, the Switch rendering is better rated than the Control rendering. We also notice that the Control and Combined renderings are not perceived with a different level of difficulty. On top of that, for the question: "I found the ability to switch renderings very useful", we obtain an average score of 6.2 +/- 0.52 (out of 7).

Ranking: Subjects were finally asked to rank the renderings in order of preference along four criteria: for identifying intersections and street corners, for identifying objects, for identifying doorways and crosswalks, and overall. Fig. 4 shows these results. We observe that the order of preference for the renderings is Switch, Combined, Control, regardless of the task performed.

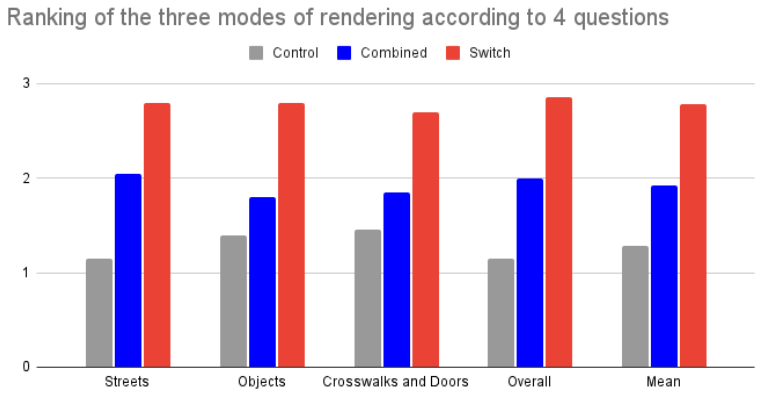

Fig. 4. Average final ranking given by users (N=20). Switch rendering is the best ranked on average, regardless of the task performed.

6 Discussion and Perspective

Our results show that the detection of crosswalks and doors is easy, regardless of the rendering used. This is not surprising because for doors, the rendering is always a white rectangle that is easy to perceive. Crosswalks are less easy to perceive than doors but

easier to perceive than other objects because they appear as a succession of parallel rectangles.

On the other hand, the identification of objects and the understanding of the street organization depend strongly on the rendering used. The results show that the Switch rendering is significantly better than the Control and Combined rendering. We observed that the Combined rendering is not significantly better than the Control rendering for object detection. This is not consistent with the study by Sanchez et al. [6]. This distinction can be explained by the resolution of the implants used in our simulator (21x18) which is lower than that used in theirs (32x32). Indeed, according to Cha et al. [14], 600 electrodes with a Scoreboard rendering (Control) are sufficient to obtain an already functional perception. The Control rendering gets more "I don't know" responses than the Switch rendering in the "objects" and "streets" tasks. The Switch rendering is systematically more used in terms of time than the Control rendering whatever the task. It is also more used than the Combined renderer in the "Doors and crosswalks" type tasks. One could imagine that with training, the decision is made more quickly with the Switch mode.

According to the analysis of subjective judgments and ranking of the rendering modes, we highlight that the Switch rendering mode has a clear advantage over the Scoreboard rendering mode in all categories. The final ranking shows that the top-ranked rendering mode is Switch over both Combined and Scoreboard.

Limitations and perspectives: However, we can address the issue of predicting the segmentation and structure images to build the different renderings. In our study, these datas are very easy to obtain because the objects of the scene are labeled. In real conditions, we would have to find another way to recover this information. The addition of neural networks would solve this problem. The prediction of semantic segmentation maps has been widely discussed lately, leading to the birth of neural networks such as EfficientDet [15] which obtain excellent performances. Regarding structure recovery there are also some networks trained to predict this kind of information such as PanoRoom [17]. There are still two problems to manage, the errors in the predictions and the calculation time. It is necessary that the predictions are not too far from reality and that these predictions are produced in quasi-real time.

Another limitation is that our study does not allow us to capture the notion of motion since we use static images. The motion information is a very useful information, so much so that some devices use event-based cameras [16] to create a phosphenic rendering. To push the realism to the maximum it could also be interesting to realize an experiment in real condition. The device could be composed of a virtual reality helmet equipped with a camera that films the scene and a smartphone that takes care of calculating the visual rendering. The interest to make a study in real conditions is double: on one hand we could propose a complete device of study very close to reality, which would allow the future works to be tested on a realistic model. Moreover, we could measure the reaction of the subjects in navigation tasks and not in perception and comprehension tasks.

References

1. Bourne, R. R. et al.: Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. The Lancet Global Health, 5(9), e888-e897 (2017)
2. Wilson, B. S. et al.: Cochlear implants: current designs and future possibilities. J Rehabil Res Dev, 45(5), 695-730. (2008)
3. Fernandez, E.: Development of visual Neuroprostheses: trends and challenges. Bio-electronic medicine, 4(1), 1-8. (2018)
4. Ratelle, A. et al.: Manuel d'intervention en orientation et mobilité. Presses de l'Université de Montréal PUM. (2019)
5. Palanker, D. et al.: Photovoltaic restoration of central vision in atrophic age-related macular degeneration. Ophthalmology, 127(8), 1097-1104. (2020)
6. Sanchez-Garcia, M.et al.: Semantic and structural image segmentation for prosthetic vision. Plos one, 15(1). (2020)
7. Zhou, D. D. et al.: The Argus II retinal prosthesis system: An overview. IEEE - (ICMEW). (2013)
8. Stingl, K. et al.: Interim results of a multicenter trial with the new electronic subretinal implant alpha AMS in 15 patients blind from inherited retinal degenerations. Frontiers in neuroscience, 11, 445. (2017)
9. Vergnieux, V. et al.: Wayfinding with Simulated Prosthetic Vision: Performance comparison with regular and structure-enhanced renderings. IEEE. (2014)
10. Vergnieux, V. et al.: Simplification of visual rendering in Simulated Prosthetic Vision facilitates navigation. Artificial organs, 41(9), 852-861. (2017)
11. Li, H.et al.: Image processing strategies based on saliency segmentation for object recognition under simulated prosthetic vision. AIM, 84, 64-78. (2018)
12. Chen, S. C. et al.: Simulating prosthetic vision: I. Visual models of phosphenes. Vision research, 49(12), 1493-1506. (2009)
13. Farvardin, M. et al.: The Argus-II retinal prosthesis implantation; from the global to local successful experience. Frontiers in neuroscience, 12, 584. (2018)
14. Cha, K. et al.: Simulation of a phosphene-based visual field: visual acuity in a pixelized vision system. Annals of biomedical engineering, 20(4), 439-449. (1992)
15. Tan, M. et al.: Efficientdet: Scalable and efficient object detection. IEEE/CVF - (CVPR). (2020)
16. Posch, C. et al.: Retinomorphic event-based vision sensors: bioinspired cameras with spiking output. Proceedings of the IEEE, 102(10), 1470-1484. (2014)
17. Fernandez-Labrador, C. et al.: Panoroom: From the sphere to the 3d layout. (2018)