Towards Photorealistic and Immersive Virtual-reality Environments for Simulated Prosthetic Vision: Integrating Recent Breakthroughs in Consumer Hardware and Software

Marc P. Zapf, *Student member, IEEE*, Paul B. Matteucci *Member, IEEE*,
Nigel H. Lovell, *Fellow, IEEE*, Steven Zheng and Gregg J. Suaning, *Senior Member, IEEE*

Abstract— **Simulated prosthetic vision (SPV) in normally sighted subjects is an established way of investigating the prospective efficacy of visual prosthesis designs in visually guided tasks such as mobility. To perform meaningful SPV mobility studies in computer-based environments, a credible representation of both the virtual scene to navigate and the experienced artificial vision has to be established. It is therefore prudent to make optimal use of existing hardware and software solutions when establishing a testing framework. The authors aimed at improving the realism and immersion of SPV by integrating state-of-the-art yet low-cost consumer technology. The feasibility of body motion tracking to control movement in photo-realistic virtual environments was evaluated in a pilot study. Five subjects were recruited and performed an obstacle avoidance and wayfinding task using either keyboard and mouse, gamepad or Kinect motion tracking. Walking speed and collisions were analyzed as basic measures for task performance. Kinect motion tracking resulted in lower performance as compared to classical input methods, yet results were more uniform across vision conditions. The chosen framework was successfully applied in a basic virtual task and is suited to realistically simulate real-world scenes under SPV in mobility research. Classical input peripherals remain a feasible and effective way of controlling the virtual movement. Motion tracking, despite its limitations and early state of implementation, is intuitive and can eliminate between-subject differences due to familiarity to established input methods.**

I. INTRODUCTION

Visual prostheses have been shown to elicit rudimentary yet useful artificial vision by eliciting a pixelized representation of the visual scene (so-called phosphenes) via stimulating electrodes placed on the visual cortex [1], the optic nerve [2] or the retina [3]. The geometry and placement of an electrode array on the retina generally directly relates to resulting acuity and field of artificial vision in a prospective recipient [4]. Manufacturing a retinal implant and subsequent testing of electrode array configurations and stimulation strategies *in vitro*, *in vivo* and in clinical trials is costly and time-consuming, spanning over several years. Therefore, effort has to be made to evaluate the prospective efficacy of electrode array configurations prior to locking in a certain design. A promising approach includes simulations of prosthetic vision (SPV) in normally-sighted volunteers. Commonly, subjects, being deprived of their normal vision, are presented with pre-processed image material resembling the vision as experienced with a hypothetical, promising prosthesis design (see Fig. 1).

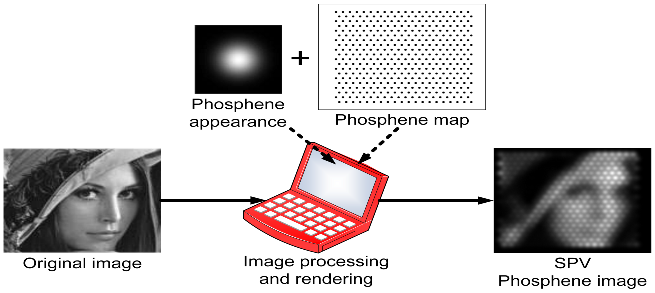

Fig. 1. Concept of SPV. The selected source image material is sampled for luminance at defined locations. Based on selected parameters for appearance of visual percepts (phosphenes) and their placement in the visual field (mapping), the image is subsequently rendered as a phosphene representation.

The visual performance of subjects in various kinds of visual tasks is subsequently observed. SPV has been recently used to investigate the efficacy of retinal prostheses in reading [5], object recognition [6] and mobility, the latter being of particular importance for persons experiencing constricting visual fields due to retinitis pigmentosa [7].

In order to obtain valid information about the mobility efficiency of a prosthesis design in SPV, both the environment the subjects are to navigate and the quality of the presented prosthetic vision have to be credible, realistic and non-restraining yet also flexible and safe. Previous work in this field involved creation of real-world as well as virtual environments [8]–[10]; however both approaches have limitations. Real-world mazes by definition provide superior immersion and body control, but can pose tripping or falling hazards and the carried simulator backpack solution restrains freedom of movement. Also, creation of extensive and dynamic environments involves great effort. Virtual mazes in SPV are a controllable and flexible alternative, yet so far have not exploited the potential of modern graphics and physics engines - obstacles and mazes were often over-simplified. A further obstacle for full immersion is use of game controllers for movement, thus lacking a sensation for movement in virtual space. Moreover, generally SPV has been restricted, due to the constraints of head-mounted displays (HMDs), to central phosphene grids of no more than 55° field of view

---

This research was supported by the Australian Research Council through its Special Research Initiative in Bionic Vision Science and Technology grant to Bionic Vision Australia.

M.P. Zapf, P. Matteucci, N.H. Lovell, S. Zheng and G.J. Suaning are with the Faculty of Engineering, Graduate School of Biomedical Engineering, UNSW Australia, Kensington, NSW, 2052, Australia. `M.zapf@unsw.edu.au`

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

(FOV), limiting research on peripheral perception.

It is therefore desirable to integrate recent hardware and software solutions in virtual environments to provide photorealism, credibility and an alternative to real-world setups. Here we present the methodology for an immersive and realistic testing framework for future SPV research, providing flexible creation, adjustment and display of mobility studies under SPV. Using a subset of the system's data analysis capability, we investigated subject speed and obstacle collisions as mobility performance measures to infer on the feasibility of different controls for virtual movement and in particular the integration of body motion capture to increase immersion.

II. METHODS

A. *Software and SPV Parameters*

The freely available CryEngine game engine (CRYENGINE 3.5.6, Crytek, Frankfurt, Germany) for creation of photorealistic virtual environments with credible physics and artificial intelligence was used to design an obstacle course. CryEngine open-source C++ code was modified and written to integrate head and body tracking to transfer real visual scanning and body motion into the virtual scene (see Hardware section). Body tracking code was adapted based on [11]. Logging of speed and timing of collisions with the environment and general timing of events was similarly integrated.

Customised SPV software for flexible real-time presentation of prosthetic vision based on a video stream (PhospheneStudio 2.0) was written by the authors in C#. 640x400 pixel resolution video was continuously captured from the CryEngine environment and, adhering to the principle depicted in Fig. 1, sampled at either 10x10 or 20x20 central hexagonal grid positions with a separation between the sampling locations of 2.45° visual angle (VA) to match the electrode spacing of the prosthesis developed by the authors (Fig. 2). A Gaussian filtering scheme with a Gaussian filter width of σ = 1.62° VA (0.66x sampling separation) was used [12]. Hexagonal arrays of 10x10 or 20x20 phosphenes with a Gaussian intensity profile (Fig. 1: Phosphene appearance) were rendered and displayed on the HMD via the freeware Deskope (Deskope 1.1 by A. Avila).

Fig. 2. Image sampling and rendering. The virtual scene (left) was luminance-sampled at 10x10 or 20x20 locations (center) and rendered as corresponding luminance-based phosphene grids (right).

B. *Hardware*

The CryEngine environment was run at 1920x1080 resolution on an Intel Core i7 with 24 GB RAM and 1GB Nvidia Geforce GTX 550 Ti graphics card. A Logitech C905 webcam was used to transfer the CryEngine frames to a second computer for SPV processing (Intel Core i7, 8 GB RAM, 1GB Geforce GTX 650 Ti). SPV was displayed on an Oculus Rift (Oculus VR Inc., Irvine, CA, USA) HMD at 1280x800 resolution. Head-tracking data including yaw, pitch and roll were acquired using the Rift in-built gyroscope, accelerometer and magnetometer. For movement control in the virtual maze, mouse and keyboard as well as an Xbox Controller and Kinect for Windows (both Microsoft, Redmond, WA, USA) over USB were used. Kinect body joint position tracking was performed to translate subject motion into walking and turning in the virtual environment. Fig. 3 illustrates the hardware and software setup.

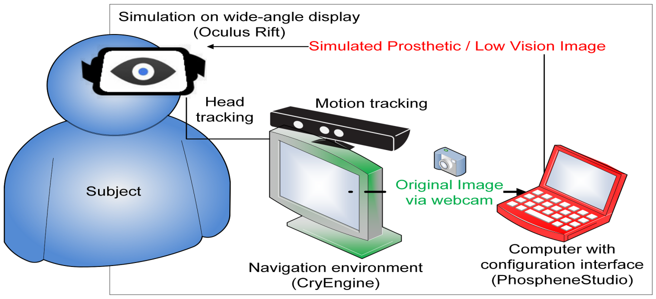

Fig. 3. Experimental setup. Subjects were instructed to navigate a CryEngine virtual scene using either keyboard/mouse, gamepad or Kinect body motion tracking. The CryEngine environment was captured, processed into SPV at 30 frames per second (fps) and displayed on the Oculus Rift HMD. Head-tracking permitted visual scanning of the scene.

C. *Motion Tracking Analysis*

An obstacle avoidance and line following task was designed and five subjects were recruited following approval by the UNSW Human Research Ethics Commitee (4 male, 1 female, mean age 25±4.8 years) to evaluate the efficiency of input methods in navigating the simulation framework. Subjects were deprived of their normal vision and provided with either the 10x10 or a 20x20 phosphene grid spanning 22° and 46.6° FOV, respectively (Fig. 4).

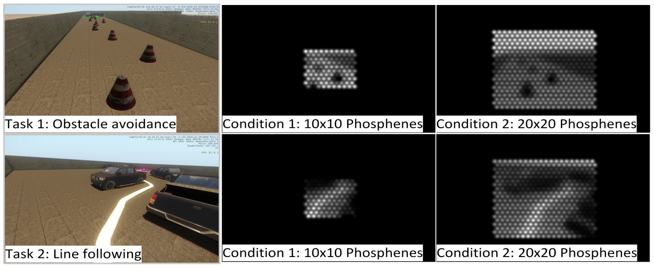

Fig. 4. Experimental conditions. Either 100 or 400 phosphenes were presented in an obstacle avoidance task (top row) and in a line following task (bottom row).

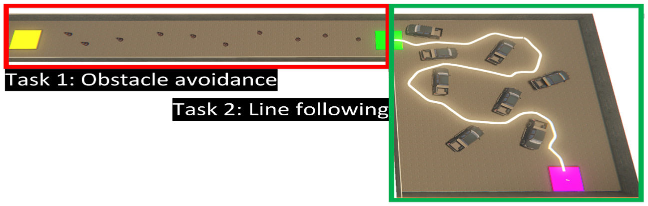

Fig. 5. Virtual obstacle course from a bird's-eye perspective.

The virtual environment consisted of two sections (Fig. 5): firstly a sidewalk with road cones as low-lying obstacles, which subjects were asked to walk around in a slalom-like fashion as fast and close as comfortable while avoiding collisions. Secondly, in a line-following task around cars as large obstacles, subjects were asked to follow the ideal line as closely as possible and as fast as comfortable. Obstacle and line positions were randomized before each run.

For each of the two vision conditions, three movement methods were evaluated: a standard mouse and keyboard setup, a gamepad and Kinect motion capture (Table 1), resulting six vision/control combinations. Using Kinect, the user was asked to face the device and turn in the virtual environment by turning the shoulders, walk sideways by leaning sideways and walk by lifting the legs on the spot in a walking-like fashion. Speed of single virtual movements could be adjusted by changing speed and extent of body movements, and two sensitivities could be chosen by a button click on a mouse in the subject's hand. Maximum walking speed was limited to 2.5 m/s in each method.

TABLE I

INPUT METHODS TO CONTROL SUBJECT MOVEMENT IN CRYENGINE

Control \ Input Method	**Keyboard + Mouse**	**Xbox Gamepad**	**Kinect Motion Tracking**
Move forward / backward	W / S	Left thumb stick forward / backward	Lift legs \ lean backward
Move sideways left / right	A / D	Left thumb stick left / right	Lean sideways
Turn left / right	Mouse movement left /right	Right thumb stick left / right	Turn shoulders left / right
Adjust speed	Mouse wheel up / down	Right thumb stick forward / backward	Adjust leg lifting frequency

The six vision/control combinations were randomized in sequence and together comprised one session. Per subject, five consecutive sessions over multiple days were run. Prior to every second session, subjects performed the task in normal vision conditions to evaluate task performance unrelated to vision deprivation. Movement speed and cumulative duration of subjects being in contact with obstacles were subsequently analyzed, representing basic measures of mobility performance using the vision and input methods provided.

III. RESULTS

A. *Movement Speed Analysis*

Median walking speeds for all control methods and vision conditions generally increased across subsequent sessions (Fig. 6). In the obstacle avoidance task, subjects, using keyboard and mouse, gained between 32.2% (full vision) to 41.9% (20x20 grid) from the first to the last session. Here, speeds reached a plateau at the maximum value of 2.5 m/s from session 3. Using the gamepad, while subjects generally moved slower initially, speed increases were more marked from 46.7% (10x10 grid) to 64.3% (20x20 grid). For the Kinect, subjects were not able to increase speed steadily across sessions; yet, speeds were 15.8% to 23.1% higher in session 5 as compared to session 1.

In the line following task, performance increases were generally more pronounced. Using keyboard and mouse, speed increased across five sessions from 32.2% (full vision) to 107.1% (20x20 grid). Using the gamepad, line following speed rose by 24.2% (20x20 grid), 25% (full vision) and 141.7% (10x10 grid). Subjects were able to increase speed with Kinect by 15.21% (full vision) to 62% (20x20 grid).

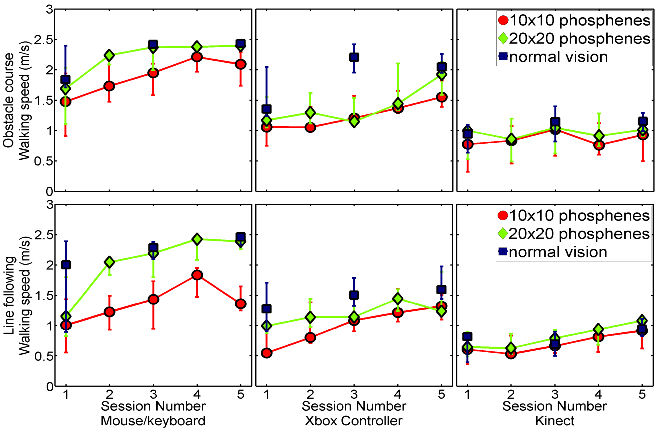

Fig. 6. Average subject walking speeds (m/s) in the obstacle avoidance (top row) and line following tasks (bottom row) for different controls and vision conditions. Squares: normal vision; circles: 10x10 phosphene grid; diamonds: 20x20 phosphene grid. Shown are median speeds with 25th&75th percentiles.

Using gamepad and Kinect, subjects started and remained at lower speeds than using the mouse and keyboard.

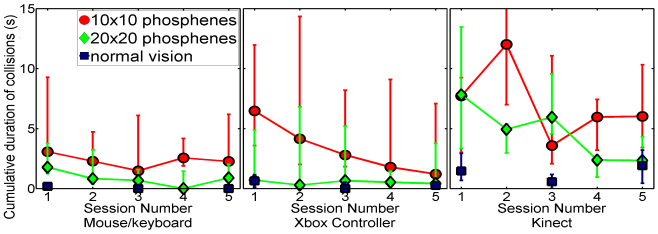

Fig. 7. Average cumulative collision durations (s) per run for the control methods and vision conditions. Squares: normal vision; circles: 10x10 phosphene grid; diamonds: 20x20 phosphene grid. Shown are median durations with 25th&75th percentiles.

B. *Collision Analysis*

Similarly to speed increases, the cumulative durations per run that subjects were in contact with obstacles and walls decreased across sessions (Fig. 7). Using mouse and keyboard, there was a 26.1% (10x10 grid) to 49.8% (20x20 grid) decrease in obstacle collision duration. For full vision, this duration approached zero. With the gamepad, when using the 10x10 grid, subjects could decrease collisions by 81.3%. With the 20x20 grid and full vision, collisions were close to zero (0.53±0.17 s and 0.3±0.3 s, respectively). Using Kinect motion capture, collisions declined by 21.9% (10x10 grid) to 70.1% (20x20 grid) from session 1 to 5, however there was an initial marked increase in collision time from session 1 to 2 (+55.7% for 10x10 grid). With full vision, collision time could not be continuously decreased.

IV. DISCUSSION AND CONCLUSIONS

In this paper we have presented a framework for mobility studies under SPV, taking advantage of recent developments on the consumer hardware market and the potential of modern gaming engines in simulating realistic virtual environments. CryEngine provides superior graphics and physics, and is regularly updated to reflect the current state of computer processing power. Despite not been used in this pilot study, CryEngine's artificial intelligence system allows presentation of credible pedestrian and traffic behaviour, situations where blinded people report having difficulties. Current PC lab equipment can run these scenarios at high frame rates of at least 30 frames per second (fps). Equally, PhospheneStudio has shown its capability to process high-resolution phosphene vision at 30 fps and permits flexible adjustment of various phosphene parameters. The results show that subjects were able to perform basic virtual tasks with an array of 100 phosphenes. It is therefore promising to apply this framework for studying the vision perceived with recent prostheses, featuring no more than 100 electrodes. It is desirable to integrate recent findings on actual phosphene shapes produced by latest clinical trials [4] to further enhance realism. Oculus Rift has been shown to be a cost-effective solution for immersion in CryEngine-based SPV. Its wide FOV of 110° horizontally, while this has not been exploited in the present study, can simulate peripheral perception. Integrated head-tracking interacted seamlessly with Kinect body tracking to provide perception of full body movement.

Subjects generally were able to increase performance by adapting to the experimental setup. Despite the lower task performance of subjects when using Kinect as opposed to classical input methods, performance nevertheless tended to increase. This demonstrates adaptation of subjects to the initially novel input method. Potentially, with further training, subjects can reach a performance comparable to the other input methods. Interestingly, movement speeds using Kinect were more coherent across different vision conditions, suggesting robustness to changing experimental conditions. An increase in Kinect performance was apparent in line following, yet not in obstacle avoidance. Task-based performance comparisons can be helpful in identifying parts of the motion tracking algorithm which have to be improved for better usability - in this case sidestepping functionality. While not translating to improved absolute performance, Kinect has been reported by subjects to provide a more realistic feeling for body movement. Previous studies demonstrating sufficient temporal and spatial accuracy of Kinect human pose estimation [13], [14] back the notion that with training, algorithm improvement and implementation of next-generation Kinect 2.0 sensors this method has the potential to more closely resemble a real-world situation and to better infer from SPV results on real-world performance.

The current pilot study included a rather young subject set. The adequacy of the presented input methods in older subjects possibly less familiar to computer technology will have to be tested. Body tracking might be more intuitive to subjects across a larger age range.

Ultimately, to prove validity of realistic SPV approaches, real-world studies will have to be conducted and compared to SPV findings. A portable version of the shown framework is in development. Apart from mobility, virtual and portable setups can be applied to other visual tasks such as object discrimination, hand-eye coordination and reading.

Creating realistic mobility studies comes with an increase in task complexity. In some conditions, there were deviations from an otherwise steady increase of performance, or a plateau was reached. Further sessions and application of further performance measures (e.g. trajectory and visual scanning analysis) will be crucial to form a comprehensive picture of the complex situations arising (losing the way, etc.) and accommodation and learning effects in mobility tasks.

Combining multiple technological advances, the presented methodology addresses several of the mentioned shortcomings of previous research. By using inexpensive or free consumer hardware and software technology which is regularly upgraded by the manufacturers, our SPV approach is expected to continuously come closer to reality. The computer industry's focus on virtual reality and alternative computer control is supporting this goal.

REFERENCES

[1] W. H. Dobelle, "Artificial vision for the blind by connecting a television camera to the visual cortex," *ASAIO journal (American Society for Artificial Internal Organs: 1992)*, vol. 46, no. 1, pp. 3–9, Feb. 2000.

[2] C. Veraart, C. Raftopoulos, J. Mortimer, J. Delbeke, D. Pins, G. Michaux, A. Vanlierde, S. Parrini, and M.-C. Wanet-Defalque, "Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode," *Brain Research*, vol. 813, no. 1, pp. 181 – 186, 1998.

[3] E. Zrenner, "Will retinal implants restore vision?" *Science*, vol. 295, no. 5557, pp. 1022–1025, 2002.

[4] R. Wilke, V.-P. Gabel, H. Sachs, K.-U. B. Schmidt, F. Gekeler, D. Besch, P. Szurman, A. Stett, B. Wilhelm, T. Peters, A. Harscher, U. Greppmaier, S. Kibbel, H. Benav, A. Bruckmann, K. Stingl, A. Kusnyerik, and E. Zrenner, "Spatial resolution and perception of patterns mediated by a subretinal 16-electrode array in patients blinded by hereditary retinal dystrophies," *IOVS*, vol. 52, no. 8, pp. 5995–6003, Jul. 2011.

[5] Y. Zhao, Y. Lu, J. Zhao, K. Wang, Q. Ren, K. Wu, and X. Chai, "Reading pixelized paragraphs of chinese characters using simulated prosthetic vision." *Invest Ophthalmol Vis Sci*, vol. 52, no. 8, pp. 5987–5994, Jul 2011.

[6] Y. Lu, J. Wang, H. Wu, L. Li, X. Cao, and X. Chai, "Recognition of objects in simulated irregular phosphene maps for an epiretinal prosthesis," *Artificial Organs*, vol. 38, no. 2, pp. E10–E20, 2014.

[7] K. A. Turano, D. R. Geruschat, J. W. Stahl, and R. W. Massof, "Perceived visual ability for independent mobility in persons with retinitis pigmentosa." *IOVS*, vol. 40, no. 5, pp. 865–877, Apr. 1999.

[8] G. Dagnelie, P. Keane, V. Narla, L. Yang, J. Weiland, and M. Humayun, "Real and virtual mobility performance in simulated prosthetic vision," *Journal of Neural Engineering*, vol. 4, no. 1, pp. S92–S101, Mar. 2007.

[9] L. Wang, L. Yang, and G. Dagnelie, "Virtual wayfinding using simulated prosthetic vision in gaze-locked viewing." *Optom Vis Sci*, vol. 85, no. 11, pp. E1057–E1063, Nov 2008.

[10] N. Parikh, L. Itti, M. Humayun, and J. Weiland, "Performance of visually guided tasks using simulated prosthetic vision and saliency-based cues." *J Neural Eng*, vol. 10, no. 2, p. 026017, Apr 2013.

[11] S. Newton, R. Lowe, R. Kember, R. Wang, and S. Davey, "The situation engine: a hyper-immersive platform for construction workplace simulation and learning," in *Proceedings of the 13th International Conference on Construction Applications of Virtual Reality*, 2013, pp. 345–353.

[12] S. Chen, N. Lovell, and G. Suaning, "Effect on prosthetic vision visual acuity by filtering schemes, filter cut-off frequency and phosphene matrix: a virtual reality simulation," in *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE*, vol. 2, Sep. 2004, pp. 4201 –4204.

[13] R. A. Clark, Y.-H. Pua, K. Fortin, C. Ritchie, K. E. Webster, L. Denehy, and A. L. Bryant, "Validity of the microsoft kinect for assessment of postural control," *Gait Posture*, vol. 36, no. 3, pp. 372–377, Jul. 2012.

[14] B. Galna, G. Barry, D. Jackson, D. Mhiripiri, P. Olivier, and L. Rochester, "Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease," *Gait Posture*, vol. 39, no. 4, pp. 1062–1068, Apr. 2014.