33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Digital Image Processing for Visual Prosthesis: Filtering Implications.

Alejandro Barriga-Rivera, *Student Member, IEEE*, and Gregg J. Suaning, *Member, IEEE*

***Abstract*— Investigators around the world are working on retinal neurostimulation as it may restore functional vision to the blind. The image is captured by a camera and after being processed, a series of electrical stimuli are applied to the surviving ganglion cells of the retina. This visual perception is expected to have low resolution. Therefore, there is a need of new algorithms that present the information contained in a visual scene understandable to humans. This study presents a novel multi-resolution algorithm based on wavelet analysis to extract the useful features of an image. Participants in this experiment were able to configure a filter bank to complete a set of everyday tasks. This study shows that wavelet-based algorithms may facilitate improved functional performance in prosthetic vision.**

I. INTRODUCTION

Restoring vision to the blind may not remain as an unsolved problem. Clinical trials have proved it is possible to induce a visual sensation by electrical stimulation [1]. As a consequence, researchers are investigating different visual neuroprosthesis prototypes to drive electrical current to the surviving neural tissue. This approach would ideally restore functional vision to patients suffering from diseases like retinitis pigmentosa or age-related macular degeneration. In particular, the stimulation of the retinal ganglion cells has been reported to produce the perception of spots of light in legally blind people [8]. These elements of light, known as phosphenes, have been the subject of different studies in the scientific literature. Several profiles have been described by Chen et al. [4]. However, the most extended model, which can be contrasted with computational analysis [6], is the Gaussian profile: a mathematical representation of the luminance of the phosphene.

Psychophysical experiments can be carried out by means of simulated prosthetic vision (SPV) [5]. A normally sighted subject is immersed in bionic vision by virtual reality. Then, a personal computer (PC) transforms the images captured by a camera and displays them using a headset. The first generation of visual prosthesis is expected to have a small number of phosphenes, big gaps between them and a limited number of brightness levels [11]. Therefore, there is a need for image processing algorithms that are able to extract the most useful information for the recipient from every scene according to the circumstance.

Mallat et al. [9] set the principles of multi-resolution analysis. This idea can be applied to image compression and to reduce its resolution. It allows one to extract the information in a multi-resolution way, copying some of the mechanisms of the eye and brain in the processing of vision. Nonetheless, context-based processing is required to facilitate functional vision: object manipulation, pattern recognition, mobility, etc.

A. Barriga-Rivera is with the Graduate School of Biomedical Engineering, University of New South Wales, NSW 2052 Australia

G. J. Suaning is with the Graduate School of Biomedical Engineering, University of New South Wales, NSW 2052 Australia

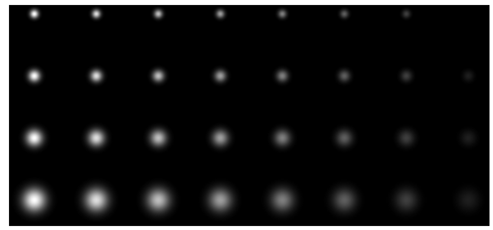

Fig. 1. The phosphene map used to render simulated prosthetic vision. Every row corresponds to a different diameter whereas columms correspond to the brightness of the phosphene.

II. METHODOLOGY

A phosphene map was rendered using 2D-Gaussian profiles. Phosphenes were considered as the elements of a vectorial base when constructing the image. Some phosphenes may overlap each other, and fill the void between adjacent phosphenes.

$$\hat{I}(x,y) = \sum_i A_i exp(-\frac{(x-x_i)^2}{\sigma_i^2}) exp(-\frac{(y-y_i)^2}{\sigma_i^2}) \quad (1)$$

$\hat{I}(x,y)$ represents the image in SPV. It is calculated as a linear combination of the elements of the phosphene map for every region of interest (i) in the original image, as described in (1). Phosphenes have been modulated in amplitude (A_i) and diameter (σ_i). In particular, 31 phosphens plus the null phosphene configure the suggested map, having eight levels of brightness (three bits) and four different diameters (two bits). To avoid pixellation, the Gaussian profile of every phosphene was represented in a 64×64 pixel array.

A hexagonal pattern of 98 phosphenes has been chosen for the phosphene lattice [7]. The distance between phosphenes varies between 0.8°and 1.2° (20 and 30 pixels respectively). Every image was first converted from a red, green, blue (RGB) image into a grayscale image (1 byte per pixel). Afterwards, it was low and high pass filtered using an orthogonal filterbank as described by Mallat et al. [9]. The low pass (LP) filter was chosen a symmetric Gaussian filter of order 10 [10].

$$h_{hp}[n] = (-1)^{(1-n)} h_{lp}[1-n] \quad (2)$$

This allows to obtain the information of different regions about the point of interest, so the phosphene would represent the LP or the high pass (HP) information of the chosen scale. In this study, participants were able to manipulate the filter algorithm. The weight of LP and HP information was balanced with the factor k as in the following equation:

$$A_i = kA_{i_{LP}} + (1-k)A_{i_{HP}} \tag{3}$$

The effect of phosphene overlapping plays an important role in the performance of the algorithm. It fills the gaps between the electrodes and provides a perception of a continuous image, what helps the brain to understand the scene. In reality, it is expected that the electrical stimulation will produce a similar perception, as the patterns of the activated excitable tissue may overlap in computational models [6].

The headset for virtual reality was provided with a camera and a display, having a resolution of 640 $\times$ 480 pixels, representing a field of view of $25° \times 19.2°$. A USB control provided with three knobs was constructed to configure the filter settings. By turning the knobs, the participant was able to vary the filter configuration, the scale factor and the phosphene representation, as described lately. The software was developed in C++/Qt under Linux.

This study was authorised by the Human Research Ethics Committee of the University of New South Wales, Australia, with reference number HREC10135. Six normally sighted participants were recruited for this experiment, three males and three females aged 18-33 years. None of them had previously participated in SPV studies. Participants A1, A2 and A3 had a phosphene spacing of 30 pixels, whereas for participants B1, B2 and B3 this parameter was configured to 20 pixels. A PC with a 14 inch screen was used as a display to show the images to the participants (they were asked to look at the screen and complete the tasks). The working area consisted of a black A2 size cardboard having an embossed white cross in its center, as a tactile and visual feedback. Four main tasks were studied in this research: motion tracking, pattern recognition, size classification and reading.

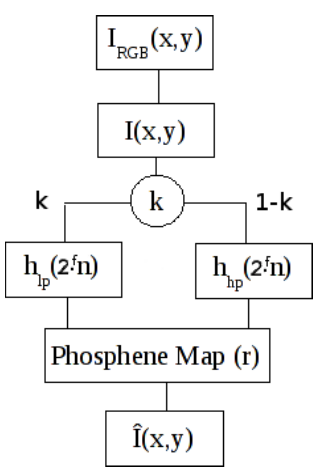

Fig. 2. The image is first converted into a gray-scale image. Afterwards, it is LP and HP filtered according to the scale factor chosen by the user (f). The coefficient corresponding to each electrode location is weighted (k) for a representation in the phosphene space, taking into account the phosphene representation factor (r).

Fig. 4. The figure shows, in the center, a plus sign as shown to the participants during pattern recognition. A high pass phosphenized image of the original is shown on the left, and the low pass version on the right.

- Motion tracking: a black screen having a small gray cross in the center was crossed randomly by a moving spot of light. Participants were required to indicate the direction of the movement (up, down, left, right).
- Pattern recognition: subjects were shown basic patterns (cross, circle, square and triangle) and the were asked to identify each of them.
- Size classification: different objects were printed in four different sizes on paper cards. Participants were required to classify their size.
- Reading: several sets of five cards containing alphanumeric characters were given to the participants. They were asked to read a sequence and to order the cards according to the sequence provided.

During the simulation, subjects were able to configure three parameters: the LP/HP ratio (k), scale factor (f) and the phosphene representation (r). The LP/HP ratio has been described previously in (3). The algorithm analyzes different vicinities around a pixel by up-sampling the filter bank [12]. The scale factor indicates how many times the filter bank is going to be up-sampled. The scaling factor narrows the bandwidth of the LP and HP filters, as indicated in the following equation, that represents the time scaling property of Fourier transform for a given signal $h[n]$.

$$h[2^f \cdot n] \;\;\Leftrightarrow\;\; 0.5H(j\Omega/2^f), \qquad f \in \mathbb{Z} \tag{4}$$

Phosphene representation allows the user to choose the phosphene size by modulating its width (σ). Once this parameter is fixed, the brightness of the phosphene is determined with 3 bits. Thus, LP/HP ratio, scale factor and phosphene representation are related to each other for an optimal representation.

III. RESULTS

The performance of each subject was assessed looking at the number of correct answers and the time required to complete the task, S and T_t respectively.

$$S = \frac{n_r - n_w/p}{n}, \qquad T_t = T(1 + \frac{n_w}{n}) \tag{5}$$

Where the total number of questions is given by n (nine sets of 20 questions per session for motion tracking and pattern recognition, and five set of 5 cards for size classification and reading), the number of right and wrong answers are n_r and n_w respectively, and p represents the number of options for each question. The time required to complete every task (T), has been adjusted accounting for the number of errors. The

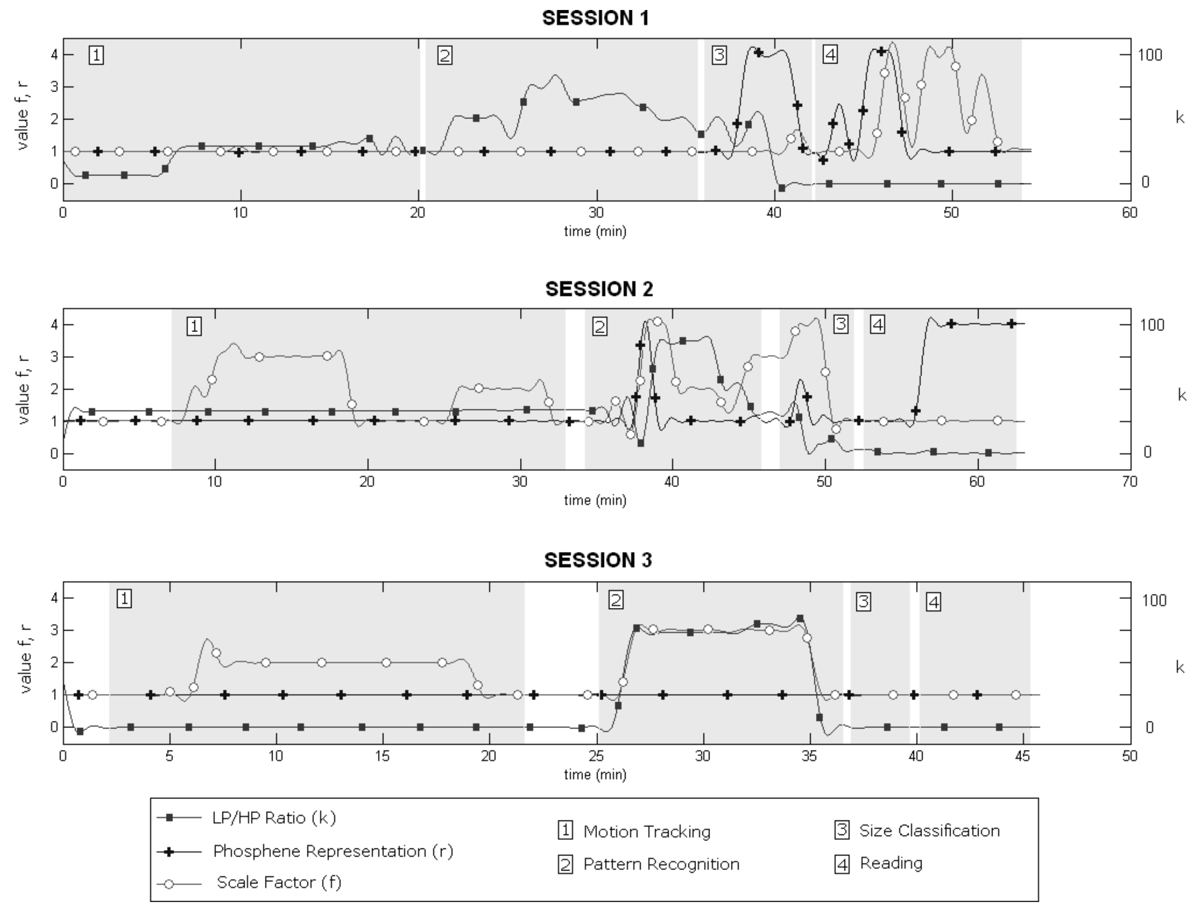

Fig. 3. The graph shows the Phosphene Representation (r) and the Scale Factor (f) on the left vertical axis, and the LP/HP Ratio (k) in percentage on the right vertical axis for participant A2. All values are represented against time in minutes and the section corresponding to each task has been shadowed and numbered as described in the legend. The scale factor looks at the size of the area analyzed around a given pixel, the phosphene representation is related to the phosphene diameter wheras the filtering ratio describes the proportion of each band that has been considered.

average value for every task can be seen in Table I for the Group A of subjects and Table II for the Group B.

Scaling factor and phosphene representation play an important role in how the algorithm extracts the information from the scene and its reconstruction. Fig. 3 represents the configuration adopted by participant A2 during the experiments. Note that both parameters are correlated to the LP/HP ratio. In particular, the correlation between the filter ratio and the scale factor for A2 was -0.3274 for session 1, 0.2957 for session 2 and 0.7699 for the last session ($p < 0.05$).

During the experiment, the filter configuration was recorded as well. It was specially interesting to observe how participants decided to use a HP or a LP filter configuration depending on their requirement for a given situation. In Fig. 3, the right vertical axis represents the factor k in percentage, as in (3). Subject A2 achieved a very good performance in size classification by setting the filtering to high pass for this task, as can be deduced from Fig. 3. The participant developed learning with each session and the behaviour became very clear in the session 3 [2]. Other subjects resorted to changing filter configurations when having problems in completing a task, giving rise to spikes in the filter configuration graph.

IV. Discussion

Several scientists have described the importance of rehabilitation programs after implantation [3], [5]. The visual cortex is expected to assist the recipient of a visual prosthesis in adapting to a phosphene-like visual perceptions. Thus, participants in this study improved their performance in every task as they were trained providing the researchers with information on how to configure a multiscale-multiresolution algorithm.

Participants having a greater phosphene to phosphene distance (Group A) achieved a lower score in general, although they showed a similar learning rate. The amount of information contained in an image is independent of the phosphene spacing, nevertheless, field of view and resolution are not: the smaller the phosphene spacing the smaller the the field of view, and therefore, the greater the resolution of a phosphenized image. In this case, considering 98 electrodes and the representation parameter fixed, an image contains 294 bits of information. Therefore, if the original image (640×480 RGB pixels) is represented with 294 bits, the compression rate is approximatelly 1:25000.

During the study, participant A3 used hands as a feedback for size classification. Subjects belonging to group A

TABLE I

RESULTS SHOWN IN GROUP A IN EACH TASK. FOR EACH SUBJECT, COLUMNS REPRESENT THE SESSION

	Subject A1			Subject A2			Subject A3		
Motion Tracking [%]	85	96.5	100	100	94.5	100	100	99	100
Pattern Recognition [seconds]	74	56	51	121	58.5	52	85	55	41.5
Size Classification [seconds]	108	74	99	70	57	51	148	210.5	185
Reading [seconds]	176	149	154	146	120.5	72.5	409	253.5	185

TABLE II

RESULTS SHOWN IN GROUP B IN EACH TASK. FOR EACH SUBJECT, COLUMNS REPRESENT THE SESSION

	Subject B1			Subject B2			Subject B3		
Motion Tracking [%]	95	98	99	92	97	95.5	94.5	98.5	99
Pattern Recognition [seconds]	118	85	49	240	59	46.5	123	65	46.5
Size Classification [seconds]	252	96.4	92.8	324	92.5	67.5	108	40.5	46.0
Reading [seconds]	180	257.5	180	300	180	129.5	119	74.5	70.5

required less time to complete the reading task. It was scored best by participant B3, even having a greater phosphene spacing. The fact that a greater phosphene spacing provides a greater field of view seems to be the cause of this result, as it involves eye-hand coordination. In addition, subjects of Group B kept a greater distance between the camera and the object they were manipulating and observing. An intermediate value of the filtering ratio provides a silhouette of the object and information about its texture. Nevertheless, none of the participants found this configuration of their interest.

V. CONCLUSIONS

Psychophysical experiments constitute the basis for experimentation when it refers to a sensory perception elicited by electrical stimulation. The recipient of a bionic eye has to be able to recognise general shapes in order to develop a functional vision. Edge detection is one of the fundamental instruments to classify objects such as cups, plates, doors or windows [3]. It is required for navigation mainly, although it is equally important in pattern and object recognition. On the other hand, face recognition and reading require a higher visual acuity, which is related to phosphene spacing and resolution. This study suggests that the algorithms based on wavelet analysis may assist recipients in different tasks by adapting the synthesis algorithm to the necessity. These algorithms are broadly used in image compression [13] and seem to be of important application in generating the stimuli to be driven to the neural tissue.

Further experiments are required to contrast and to validate different configurations for particular tasks. There is a knowledge gap about the significance of band pass analysis in prosthetic vision. Ideally, it would be of relevant interest in designing a smart system able to adapt the configuration of the visual prosthesis as a guide dog decides what to do for a blind.

REFERENCES

[1] GS Brindley and WS Lewin. The sensations produced by electrical stimulation of the visual cortex. *The Journal of Physiology*, 196(2):479, 1968.

[2] SC Chen, LE Hallum, NH Lovell, and GJ Suaning. Learning prosthetic vision: a virtual-reality study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(3):249–255, 2005.

[3] SC Chen, GJ Suaning, JW Morley, and NH Lovell. Rehabilitation regimes based upon psychophysical studies of prosthetic vision. *Journal of Neural Engineering*, 6:035009, 2009.

[4] S.C. Chen, G.J. Suaning, J.W. Morley, and N.H. Lovell. Simulating prosthetic vision: I. Visual models of phosphenes. *Vision research*, 49(12):1493–1506, 2009.

[5] G. Dagnelie. Psychophysical evaluation for visual prosthesis. 2008.

[6] S. Dokos, GJ Suaning, and NH Lovell. A bidomain model of epiretinal stimulation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(2):137–146, 2005.

[7] NB Dommel, YT Wong, T. Lehmann, CW Dodds, NH Lovell, and GJ Suaning. A CMOS retinal neurostimulator capable of focussed, simultaneous stimulation. *Journal of Neural Engineering*, 6:035006, 2009.

[8] M.S. Humayun, E. de Juan Jr, G. Dagnelie, R.J. Greenberg, R.H. Propst, and D.H. Phillips. Visual perception elicited by electrical stimulation of retina in blind humans. *Archives of Ophthalmology*, 114(1):40, 1996.

[9] S.G. Mallat et al. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989.

[10] R. Navarro and A. Tabernero. Gaussian wavelet transform: two alternative fast implementations for images. *Multidimensional Systems and Signal Processing*, 2(4):421–436, 1991.

[11] N.R. Srivastava, P.R. Troyk, and G. Dagnelie. Detection, eye–hand coordination and virtual mobility performance in simulated vision. *Journal of Neural Engineering*, 6:035008, 2009.

[12] M. Vetterli and C. Herley. Wavelets and filter banks: Theory and design. *IEEE Transactions on Signal Processing*, 40(9):2207–2232, 1992.

[13] J.D. Villasenor, B. Belzer, and J. Liao. Wavelet filter evaluation for image compression. *IEEE Transactions on image processing*, 4(8):1053–1060, 1995.