31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Neurophysiology Study of Early Visual Processing of Face and Non-face Recognition under Simulated Prosthetic Vision

Yuan Yang, Hong Guo, Shanbao Tong, *Member, IEEE*, Yisheng Zhu, *Senior Member, IEEE* and Yihong Qiu, *Member, IEEE*

Abstract—**Behavioral researches have shown that the visual function can be partly restored by phosphene-based prosthetic vision for the non-congenital blinds. However, the early visual processing mechanisms of phosphene object recognition is still unclear. This paper aimed to investigate the electro-neurophysiology underlying the phosphene face and non-face recognition. The modulations of latency and amplitude of N170 component in the event-related potential (ERP) were analyzed. Our preliminary results showed that (1) both normal and phosphene face stimuli could elicit prominent N170; nevertheless, phosphene stimuli caused notable latency delay and amplitude suppression on N170 compared with normal stimuli and (2) under phosphene non-face stimuli, a slight but significant latency delay occurred compared with normal stimuli, while amplitude suppression was not observed. Therefore, it was suggested that (1) phosphene perception caused a disruption of the early visual processing for non-canonical images of objects, which was more profound in phosphene face processing; (2) the face-specific processing was reserved under prosthetic vision and (3) holistic processing was the major stage in early visual processing of phosphene face recognition, while part-based processing was attenuated due to the loss of the details.**

I. Introduction

Several research groups in the world have developed prototype visual prosthetic systems to partly restore the visual function by electrically stimulating the visual pathway of the non-congenital blinds [1-3]. Behavioral studies have validated the effectiveness of phosphene-based prosthetic vision in pattern recognition [4], text reading [5, 6] and facial recognition [7]. According to Humphreys [8], the phosphene object recognition employed two procedures: (1) the early visual processing of phosphene patterns, i.e., object perception, and (2) the late procedure of associating perceived information with the models in the brain, i.e., object identification. From the view of cognitive psychology, the early visual processing was a fundamental phase in object recognition, particularly for an unknown phosphene object [9]. However, the neurophysiologic mechanisms of the early visual processing of phosphene objects were still unclear. Therefore, this study aimed to investigate the early visual processing mechanisms of object recognition under simulated prosthetic vision.

Event-related potential (ERP) study has been widely used for investigating the neurophysiologic mechanisms of object recognition. Neurophysiologic results on face perception have reported a negative-going potential in ERPs occurring between 170 and 200 ms after stimuli onset (N170) at occipito-temporal electrodes [10, 11], which was relatively larger for faces than that for non-face objects [12]. Some studies suggested the N170 component be associated with structural encoding process of face recognition in occipital-temporal areas [13]. Also isolated eyes, filtered faces or inverted face could elicit N170 with delayed latencies and/or modulated amplitudes, which was attributed to two mechanisms of early visual processing, i.e., holistic processing and part-based processing [14-16]. Therefore, we designed an object perception experiment using photographs of normal faces, normal non-face objects and their phosphene versions, during which electroencephalogram (EEG) was recorded with scalp electrodes. The amplitudes and latencies of N170 components were analyzed.

II. Methods

A. Subjects

Seven right-handed undergraduate students (female/male: 2/5; age: 22 ± 4) with normal or corrected-to-normal vision participated in the experiments. All subjects signed a written informed consent and were compensated for their participations.

B. Stimuli and Experimental Design

Four categories of photographs (189 × 224 pixels) included 10 human faces (female/male: 5/5), 10 buildings, 10 chairs and 10 birds (target stimuli). All face photographs were selected from 'The Database of Faces' at the AT&T Laboratories Cambridge (formerly 'The ORL Database of faces') [17]. A mask of 32 × 27 grids was applied to all

This work is supported by the National Key Basic Research Program (Grant No. 2005CB724303).

Y. Qiu is with Department of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (phone: +86-21-34204717; fax: 86-21-34204717; e-mail: yhqiu@sjtu.edu.cn).

S. Tong is with Med-X Research Institute, Shanghai Jiao Tong University, Shanghai 200030, China (phone: +86-21-34205138; e-mail: shanbao.tong@gmail.com).

Y. Yang, H. Guo, Y. Zhu are with Department of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, 200240 China.

photographs except the targets to generate the corresponding phosphene stimuli, which could be sufficiently recognized by the subjects [18]. In order to maintain the same spectral energy of the stimulation, the background of all stimuli was tuned to a uniform gray level (Fig. 1).

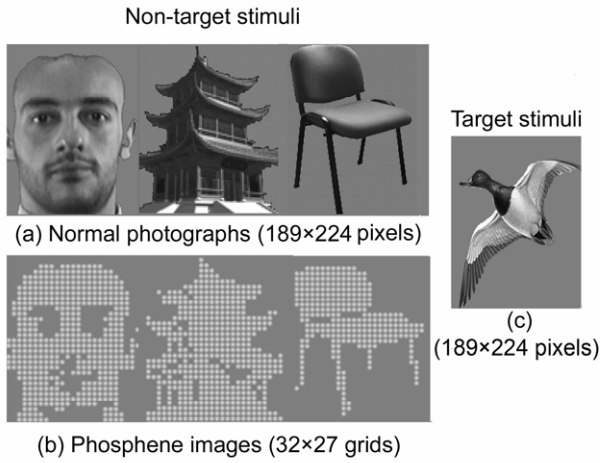

Fig. 1. The non-target stimuli included (a) normal photographs (faces, buildings and chairs) and (b) their corresponding phosphene versions. (c) Target stimuli (birds) were used as attractors in the experiment.

Participants were seated comfortably in a dark sound-and-RF shielding room (Model: Union-Brother, Guangzhou, China). Stimuli included normal images, their corresponding phosphene versions and targets, with each one repeated 12 times ($30 \times 12 \times 2 + 10 \times 12 \times 1 = 840$ trials in total). Each stimulus (6.39° × 5.22° view angle) was centered on a black screen (Model: LXM-L19CH, Lenovo, China, resolution: 800 × 600, refresh rate: 60 Hz) and randomly presented for 500 ms with a variable inter-stimulus interval (ISI: 800-1300 ms). The entire procedure was divided into ten blocks with 84 stimuli each with 5-10 min in-between breaks. In each block, the number of the target stimuli ranged from 8-16 randomly. The participants were required to count the number of target stimuli (i.e., birds in this study) in mind and to report the result at the end of each block.

C. ERP Recordings and Data Analysis

The EEG was recorded (NeuroScan 4.3, Compumedics NeuroScan, USA) via 64 electrodes (Sintered Ag/AgCl electrodes) mounted on a cap (Quik-Cap, Compumedics NeuroScan, USA) with reference to mastoid (M1&M2). Four ocular electrodes were used to monitor the vertical and horizontal eye movements at the outer canthi and left supraorbital/infraorbital ridges. Electrode impedances were kept under 5 KΩ. The sampling rate was 1000 Hz. The signals were off-line analyzed with EEGLAB 6.01v [19]. The raw EEG signal was first filtered by a 0.1-50Hz zero phase shift filter and an average reference was calculated for all electrodes. After EOG correction and EEG artifact rejection (≥ 100 μV), EEG was partitioned into 1200ms epochs with a 200 ms pre-stimulus baseline plus 1000ms post-stimulus onset. ERPs were derived by grand averaging the trials corresponding to normal faces, normal buildings, normal chairs and their phosphene versions, respectively, before being digitally filtered into 0.1-30 Hz. Since the targets were only designed as attractors for the subjects, the ERP under target stimuli (birds) was not investigated.

The amplitudes and latencies of occipito-temporal N170 components were measured at corresponding electrode sites, i.e., two pairs of posterior sites (P7/P8, PO7/PO8). In order to minimize the influences of the earlier component of P100, the amplitudes of N170 were measured as the difference between the valley of N170 and the peak of P100 [14, 20]. Analysis of variance (ANOVAs) with repeated measures and paired *t*-tests were performed to check the statistic significance of the results, and Greenhouse-Geisser corrections were made when needed. Within-participant factors included category (chair vs. building vs. human face), stimulus format (normal stimuli vs. phosphene stimuli), site (PO7/PO8 vs. P7/P8) and hemisphere (left vs. right). The dependent variables were the N170 amplitude and its latency.

III. ERP RESULTS

All subjects reported correctly the number of target stimuli. Fig. 2 showed the averaged ERPs over all the subjects at PO7 and PO8 in response to normal and phosphene faces (Fig. 2a & 2b), normal and phosphene buildings (Fig. 2c & 2d) and normal and phosphene chairs (Fig. 2e & 2f).

A. N170 Latency

The overall ANOVA showed that stimulus format had a main effect ($F(1;6) = 9.425$, $P = 0.022$) on N170 latencies, where the N170 components were significantly delayed under phosphene stimuli compared with those under normal stimuli. Moreover, phosphene faces caused a longer latency delay (14 ± 4 ms) than phosphene non-faces (building: 7 ± 1 ms, chair: 3 ± 1 ms) ($P < 0.001$). While the factors of category ($F(2;12) = 1.637$, $P = 0.235$), site ($F(1;6) = 3.776$, $P = 0.100$) and hemisphere ($F(1;6) = 0.566$, $P = 0.480$) did not show any main effects on the N170 latencies.

B. N170 Amplitude

The ANOVA revealed significant main effect of category ($F(2;12) = 40.340$, $P < 0.001$) on N170 amplitudes and interaction between category and format ($F(2;12) = 8.821$, $P < 0.004$). Both normal (5.5 ± 0.5 μV) and phosphene faces (5.4 ± 1.5 μV) elicited larger N170 than non-face stimuli (normal buildings: 2.1 ± 0.1 μV, phosphene buildings: 2.8 ± 0.1 μV, normal chairs: 3.4 ± 0.5μV and phosphene chairs: 4.0 ± 1.0 μV) (all $P < 0.05$). For faces, paired *t*-test showed that N170 was significantly diminished under the phosphene stimuli compared with normal stimuli ($P < 0.001$). However, this effect was not significant for non-faces (buildings: $P = 0.872$, chairs: $P = 0.573$). Main effects were also found on site ($F(1; 6) = 11.571$, $P = 0.014$) and hemisphere ($F(1; 6) = 6.277$, $P = 0.046$). Paired *t*-test verified that the amplitudes on the right hemisphere (P8/PO8) were significantly larger than

those on the left (P7/PO7) for normal faces (P = 0.008), phosphene faces (P = 0.014), normal buildings (P = 0.007), phosphene buildings (P = 0.024) and normal chairs (P = 0.013), but not for phosphene chairs (P = 0.113). Moreover, interactions of category × site (F(1.022; 6.129) = 0.492, P = 0.623), category × hemisphere (F(1.146; 6.875) = 1.924, P = 0.212), format × site (F(1; 6) = 4.310, P = 0.083), format × hemisphere (F(1; 6) = 0.287, P = 0.611) and site × hemisphere (F(1; 6) = 3.814, P = 0.099) were not significant.

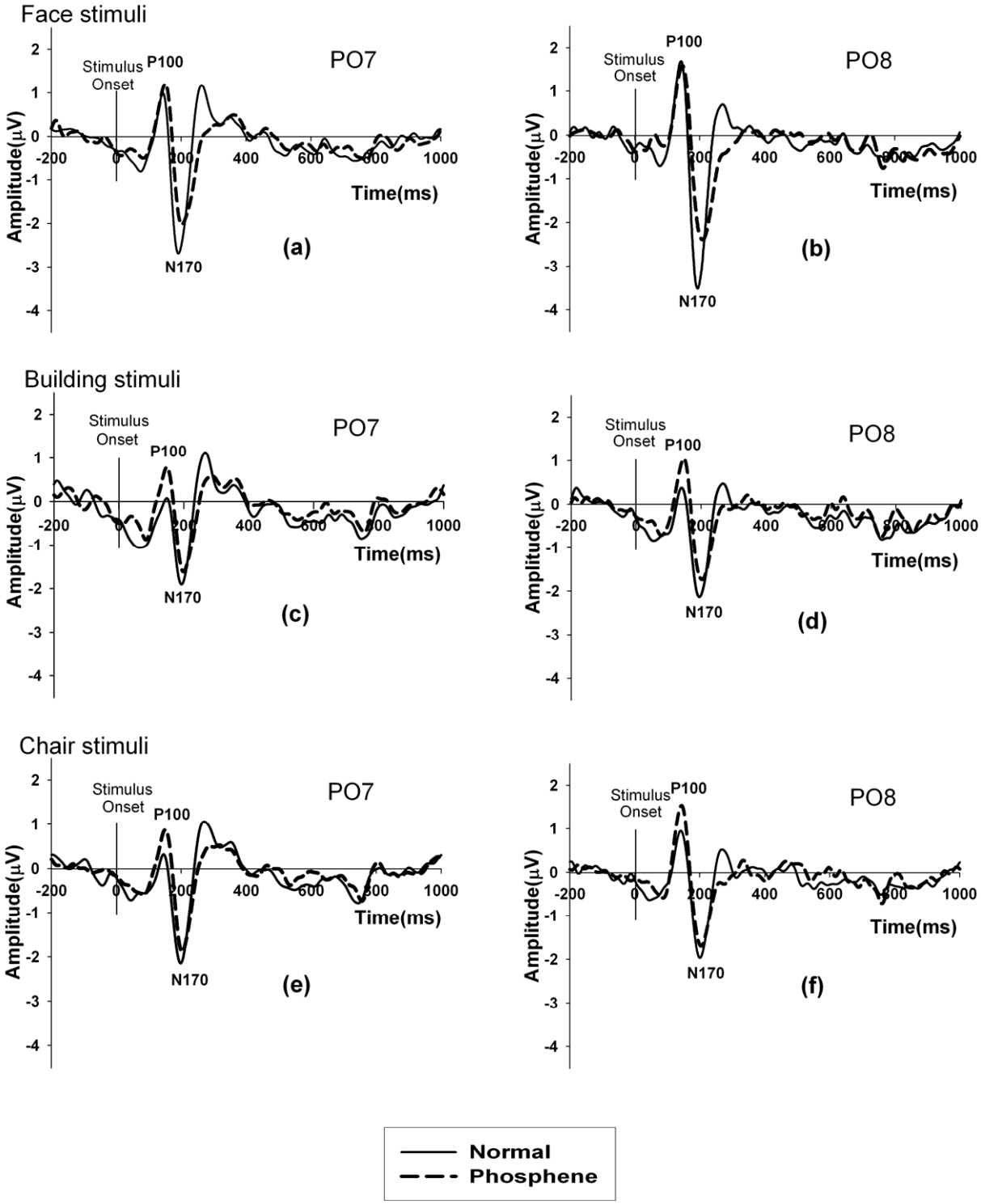

Fig. 2. Averaged ERP waveforms over all the subjects for normal/phosphene faces (a/b), normal/phosphene buildings (c/d) and normal/phosphene chairs (e/f) at PO7 and PO8, respectively.

IV. Discussion

N170 component was used to investigate the early visual processing mechanisms of object recognition with stimulated prosthetic vision. Although the simulated phosphene objects were idealized prototypes with evenly aligned phosphenes and adequate phosphene numbers, useful insights on phosphene perception mechanisms could be discovered based on such simplified patterns.

Firstly, we found that N170 was delayed by phosphene format for all categories tested. Such effect was far larger for faces than for any other categories. Similar phenomena were found by Itier *et al.* that N170 was delayed by the inverted stimuli [15], which implied a disruption of the processing for non-canonical images of objects. Since the phosphene image was also a kind of non-canonical image, the phosphene effect on N170 latencies found in our experiment confirmed such implication. Furthermore, such disruption of processing seemed to be more notable in phosphene face processing than phosphene non-face processing.

Secondly, prominent N170 was elicited by both normal and phosphene faces, which indicated that face structural encoding mechanism occurred not only under normal vision but also under prosthetic vision. As it had been discussed in a handful of previous studies, this face structural encoding mechanism might involve a complex system composed of at least two separate channels, i.e., a holistic processor and a part-based processor, which both modulated the N170 [16]. In our experiments, diminished N170 was observed under phosphene face stimuli. Noticing the loss of details under phosphene face stimuli compared to normal face stimuli, we speculated that the decline of N170 amplitude was due to the suppressed part-based processing. Moreover, a slight decrease of N170 amplitude was also discovered by Goffaux *et al.* under low-pass filtering face stimuli compared with broad-pass filtering face stimuli [14], which was similar to our results because the absence of high-frequency details in phosphene stimuli could lead to the attenuation of part-based processor. However, the N170 amplitudes were comparable under both formats of non-face stimuli, which might imply that similar early visual processing was involved in non-face recognition under both stimulus formats. Furthermore, it was also proved that face and non-face perception processes were different under prosthetic vision just as they were under normal vision.

In conclusion, our experimental results suggested that: (1) phosphene effect on latency of N170 implied a disruption of the processing for non-canonical images, which was more profound in phosphene face processing; (2) the face-specific processing was reserved under prosthetic vision and (3) holistic processing played a dominating role in phosphene face recognition while part-based processing was attenuated due to the loss of details.

Acknowledgment

The authors thank Prof. X. Zhu for providing EEG recording facilities for the experiment in the study.

References

[1] P. H. Schiller and E. J. Tehovnik, "Visual prosthesis," *Perception*, vol. 37, pp. 1529-1559, 2008.

[2] S. C. Chen, L. E. Hallum, N. H. Lovell, and G. J. Suaning, "Visual acuity measurement of prosthetic vision: a virtual-reality simulation study," *J. Neural Eng.*, vol. 2, pp. S135-S145, 2005.

[3] G. J. Suaning, L. E. Hallum, S. C. Chen, P. J. Preston, and N. H. Lovell, "Phosphene vision: development of a portable visual prosthesis system for the blind," presented at Eng. in Med. and Bio. Soc., 2003. Proc. 25th Annu. Int. Conf. IEEE, 2003.

[4] C. Veraart, M. C. Wanet-Defalque, B. Gerard, A. Vanlierde, and J. Delbeke, "Pattern recognition with the optic nerve visual prosthesis," *Artif. Organs*, vol. 27, pp. 996-1004, 2003.

[5] K. Cha, K. W. Horch, R. A. Normann, and D. K. Boman, "Reading speed with a pixelized vision system," *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.*, vol. 9, pp. 673-677, 1992.

[6] G. Dagnelie, D. Barnett, M. S. Humayun, and R. W. Thompson Jr, "Paragraph text reading using a pixelized prosthetic vision simulator: parameter dependence and task learning in free-viewing conditions," *Invest. Ophthalmol. Vis. Sci.*, vol. 47, pp. 1241-1250, 2006.

[7] R. W. Thompson, G. D. Barnett, M. S. Humayun, and G. Dagnelie, "Facial recognition using simulated prosthetic pixelized vision," *Invest. Ophthalmol. Vis. Sci.*, vol. 44, pp. 5035-5042, 2003.

[8] G. W. Humphreys, "Top-down processes in object identification: Evidence from experimental psychology, neuropsychology and functional anatomy," *Philos. Trans. R. Soc. B-Biol. Sci,* vol. 352, pp. 1275-1282, 1997.

[9] A. Delorme, G. A. Rousselet, M. J. M. Macé, and M. Fabre-Thorpe, "Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes," *Cognit. Brain Res.*, vol. 19, pp. 103-113, 2004.

[10] T. Allison, A. Puce, D. D. Spencer, and G. McCarthy, "Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli," *Cereb. Cortex*, vol. 9, pp. 415-430, 1999.

[11] S. Bentin, T. Allison, A. Puce, E. Perez, and G. McCarthy, "Electrophysiological studies of face perception in human," *J. Cogn. Neurosci.*, vol. 8, pp. 551-565, 1996.

[12] G. McCarthy, A. Puce, A. Belger, and T. Allison, "Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex," *Cereb. Cortex*, vol. 9, pp. 431-444, 1999.

[13] M. Eimer, "The face-specific N170 component reflects late stages in the structural encoding of faces," *Neuroreport*, vol. 11, pp. 2319, 2000.

[14] V. Goffaux, I. Gauthier, and B. Rossion, "Spatial scale contribution to early visual differences between face and object processing," *Cognit. Brain Res.*, vol. 16, pp. 416-424, 2003.

[15] R. J. Itier, M. Latinus, and M. J. Taylor, "Face, eye and object early processing: What is the face specificity?," *Neuroimage*, vol. 29, pp. 667-676, 2006.

[16] N. Sagiv and S. Bentin, "Structural encoding of human and schematic faces: holistic and part-based processes," *J. Cogn. Neurosci.*, vol. 13, pp. 937-951, 2001.

[17] ORL face database. AT&T Laboratories Cambridge, 1994. Available: http://www.cam-orl.co.uk/facedatabase.html. December 12, 2007.

[18] J. A. Dowling, A. J. Maeder, and W. W. Boles, "Intelligent image processing constraints for blind mobility facilitated through artificial vision," presented at Proc. 8th Aust. and N. Z. Intell. Inform. Sys. Conf. (ANZIIS), Sydney, Australian, 2003.

[19] M. Honal and T. Schultz, "Identifying user state using electroencephalographic data," presented at Proc. Int. Conf. on Multimodal Input (ICMI), Trento, Italy, 2005.

[20] S. Bentin and L. Y. Deounell, "Structural encoding and identification in face processing: ERP evidence for separate mechanisms," *Cogn. Neuropsychol.*, vol. 17, pp. 35-54, 2000.