Vision Research 137 (2017) 61–79

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Face identity recognition in simulated prosthetic vision is poorer than previously reported and can be improved by caricaturing

Jessica L. Irons [a,b], Tamara Gradden [a], Angel Zhang [a], Xuming He [c,d,f], Nick Barnes [c,d,e,f], Adele F. Scott [c,e,f], Elinor McKone [a,b,*]

[a] *Research School of Psychology, Australian National University, Australia*
[b] *ARC Centre for Cognition and Its Disorders, Australian National University, Australia*
[c] *National Information and Communication Technology Australia (NICTA), Australia*
[d] *College of Engineering and Computer Science, Australian National University, Australia*
[e] *Bionic Vision Australia, Australia*
[f] *Data61, CSIRO, Australia*

A R T I C L E I N F O

Article history:
Received 28 January 2017
Received in revised form 15 April 2017
Accepted 8 June 2017

Number of reviewer = 1

Keywords:
Prosthetic vision
Retinal prosthesis
Face recognition
Caricaturing

A B S T R A C T

The visual prosthesis (or “bionic eye”) has become a reality but provides a low resolution view of the world. Simulating prosthetic vision in normal-vision observers, previous studies report good face recognition ability using tasks that allow recognition to be achieved on the basis of information that survives low resolution well, including basic category (sex, age) and extra-face information (hairstyle, glasses). Here, we test within-category individuation for face-only information (e.g., distinguishing between multiple Caucasian young men with hair covered). Under these conditions, recognition was poor (although above chance) even for a simulated 40 × 40 array with all phosphene elements assumed functional, a resolution above the upper end of current-generation prosthetic implants. This indicates that a significant challenge is to develop methods to improve face identity recognition. Inspired by “bionic ear” improvements achieved by altering signal input to match high-level perceptual (speech) requirements, we test a high-level perceptual enhancement of face images, namely face caricaturing (exaggerating identity information away from an average face). Results show caricaturing improved identity recognition in memory and/or perception (degree by which two faces look dissimilar) down to a resolution of 32 × 32 with 30% phosphene dropout. Findings imply caricaturing may offer benefits for patients at resolutions realistic for some current-generation or in-development implants.

1. Introduction

Visual prostheses can restore partial vision to individuals blinded by conditions such as retinitis pigmentosa, by bypassing the damaged photoreceptors and electrically stimulating intact neurons. Implants can target various regions including cortex (Brindley & Lewin, 1968), optic nerve (Delbeke, Oozeer, & Veraart, 2003), and retina (Humayun et al., 2012; Stingl et al., 2013). Retinal prostheses, for example, can comprise an internal photodiode array which responds directly to incoming light (e.g., Retinal Implant AG subretinal device; Zrenner et al., 2011) or an internal microelectrode array that receives wireless input from an external image capturing system such as a camera placed on glasses worn by the patient (e.g., Second Sight’s Argus II, Humayun et al., 2012, and Bionic Vision Australia’s epiretinal and suprachoroidal devices, Ayton et al., 2014; Tran et al., 2011).

The resolution of current prosthetic devices is far below that of natural vision (Stingl et al., 2013). As illustrated in Fig. 1, the stimulation of intact neurons creates a percept of ‘phosphenes’ (balls of light) (Brindley & Lewin, 1968; Dobelle, Mladejovsky, & Girvin, 1974). For devices currently implanted in patients, electrode arrays vary in resolution, including a 6 × 10 array (Humayun et al., 2012) and a 38 × 40 array (Zrenner et al., 2011), and the operational resolution will typically be lower than the number of electrode elements, due to some electrodes not working or being implanted over dead tissue (henceforth referred to as ‘electrode dropout’). To date, no-one has shown more phosphenes than electrodes, so more electrodes can result in higher resolution, and the number of electrodes is currently the upper limit (although note that acuity

---

[*] Corresponding author at: Research School of Psychology, Australian National University, Canberra, ACT 0200, Australia.

E-mail addresses: jesirons@gmail.com (J.L. Irons), Elinor.McKone@anu.edu.au (E. McKone).

http://dx.doi.org/10.1016/j.visres.2017.06.002

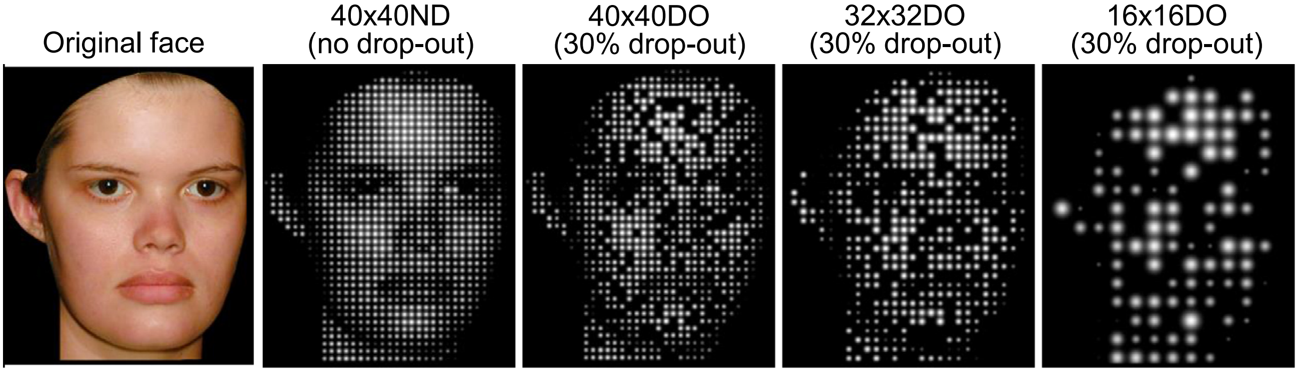

Fig. 1. Simulation of phosphene appearance in a bionic eye. Examples illustrate a single female face in full-resolution color image, followed by phosphenized versions of the same face at four resolutions.

does not necessarily scale with the number of electrodes, e.g., Humayun et al., 2012 vs Zrenner et al., 2011).

In terms of functional performance, simulations in normal-vision observers, and some patient studies, have reported that lower resolution electrode arrays (e.g., a 6 × 10 array) can be sufficient for some tasks, such as wayfinding (i.e., walking a specified route through an environment while avoiding obstacles; Barnes et al., 2015; van Rheede, Kennard, & Hicks, 2010) and object localisation (Humayun et al., 2012). Recognizing the basic category of common items (e.g., as a bicycle, shoe, or chair) appears to require at least 16 × 16, and recognizing scenes (e.g., bedroom, dining room or stairs) at least 32 × 32 (Zhao et al., 2010).

1.1. Previous studies of face recognition

Several previous studies have also tested face recognition, all using a simulation of prosthetic vision in normal-vision observers (i.e., showing observers images similar to those in Fig. 1). Results

Table 1
Accuracy of face recognition in previous studies, and the category and extra-face cues available that could have assisted performance; present study results included for comparison.

Article	Task	Category and extra-face cues available	Learning format	Scanning or snapshot?	Resolution[1]	Correct[2] (%)	Chance (%)
Thompson et al. (2003)	*Short-term memory*: Learn four unfamiliar faces. One of these (test face) shown immediately afterwards. Select which of the 4 it matched	Age, hairstyle, glasses	High resolution	Scanning	32 × 32DO 25 × 25DO 16 × 16DO	76 79 74	25 ” ”
Li et al. (2005)	”	Age, hairstyle, expression, facial hair	High resolution	Single snapshot	32 × 32 25 × 25 16 × 16	70 65 60	25 ” ”
Vurro et al. (2006)	*Simultaneous perception*: View four unfamiliar faces. Test face presented at the same time. Select which of the 4 it matches	Sex, age, hairstyle	Phos-phenized	Single snapshot	10 × 10	75	25
Chang et al. (2010)	*Long-term memory*: Identify familiar individual (colleague)	Sex, age, hairstyle, glasses	High resolution	Single snapshot	16 × 16 12 × 12 8 × 8	74 55 38	7.1 ” ”
Chang et al. (2012)	”	Sex, age, hairstyle, glasses	High resolution	Single snapshot	16 × 16 12 × 12 8 × 8	84 56 27	6.7 ” ”
Wang et al. (2014) (head[3])	”	Sex, age, hairstyle, glasses	High resolution	Single snapshot	32 × 32 24 × 24	88 65	5.6 ”
Wang et al. (2014) (face-only[3])	As above, but with hair removed (& phosphene grid tiling face not whole head)	Sex, age, glasses	High resolution	Single snapshot	32 × 32 24 × 24	77 53	” ”
Present study[4]	*Long-term memory*: old-new recognition	None	High resolution	Single snapshot	40 × 40 40 × 40DO 32 × 32DO 16 × 16DO	59 51 50 50	50 ” ” ”
			Phos-phenized	Scanning	40 × 40DO	58	”

Notes: 1. Resolutions ending in “DO” had 30% electrode dropout. All others had no dropout. For Thompson, data reported are for condition where phosphene grid tiles full head, for closest match to other studies' resolution values. 2. Accuracy averaged over all conditions reported in the article. Across studies, these varied in factors such as image size, phosphene contrast level, phosphene grid shape (e.g., rectangular vs hexagonal) and, in Chang et al. (2010, 2012) and Li et al. (2005), whether low-level image enhancement techniques (e.g., edge detection, contrast enhancement) were included. 3. The “Face-only” condition in Wang et al. (2014) refers to the VJFR-ROI condition, in which a face detection algorithm was used to zoom in on the internal facial features, cropping out most of the hair. “Head” refers to all other conditions, in which the full head including hair was visible. 4. For present study, data are for Veridical faces average accuracy for Old and New trials. Note accuracy for high-resolution test faces was 88% correct, demonstrating that the poor performance for phosphenised faces was not due to failure to learn or remember the faces themselves.