Vision Research 49 (2009) 368–373

Contents lists available at ScienceDirect

Vision Research

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

Transfer between pose and expression training in face recognition

Wenfeng Chen [a,*], Chang Hong Liu [b,1]

[a] Institute of Psychology, Chinese Academy of Sciences, 4A Datun Road, Chaoyang District, Beijing 100101, China
[b] Department of Psychology, University of Hull, Cottingham Road, Hull HU6 7RX, United Kingdom

ARTICLE INFO

Article history:
Received 2 April 2008
Received in revised form 2 September 2008

Keywords:
Face recognition
Transfer
Pose
Expression

ABSTRACT

Prior research has shown that recognition of unfamiliar faces is susceptible to image variations due to pose and expression changes. However, little is known about how these variations on a new face are learnt and handled. We aimed to investigate whether exposures to one type of variation facilitate recognition in the untrained variation. In Experiment 1, faces were trained in multiple or single pose but were tested with a new expression. In Experiment 2, faces were trained in multiple or single expression but were tested in a new pose. We found that higher level of exposure to pose information facilitated recognition of the trained face in a new expression. However, multiple-expression training failed to transfer to a new pose. The findings suggest that generalisation of pose training may be extended to different types of variation whereas generalisation of expression training is largely confined within the trained type of variation.

1. Introduction

The human face transmits rich information such as identity and expressions through a non-rigid 3D shape. Face recognition requires detection of invariant properties of this shape and its reflectance across rigid and non-rigid movements. Despite its complexity in two-dimensional image transformations, our ability to recognise faces often seems surprisingly effortless. However, there is now substantial evidence that recognition is rather fallible for unfamiliar faces (Hancock, Bruce, & Burton, 2000). Bruce (1982) demonstrated that recognition performance can be adversely impaired if a face is learned and tested in different poses or expressions. Since recognition of familiar faces is largely pose and expression invariant, this evidence suggests that image-invariant recognition requires learning or familiarisation. However, exactly how the brain learns to tackle image variations due to pose and expression remains little known.

Brain imaging research has revealed that the fusiform gyrus, a face selective area of the brain, is insensitive to low-level image variations such as size (Kanwisher, McDermott, & Chun, 1997). However, adaptation in this area to repeated presentations of a face is sensitive to variations of pose and facial expressions (Andrews & Ewbank, 2004). Because activities of some face selective neurons are tuned to specific poses (Abbott, Rolls, & Tovee, 1996), there has been a suggestion that pose-invariant representations are formed by converging information from pose-dependent neurons (Booth & Rolls, 1998).

Psychophysical research has also produced evidence for pose-dependent recognition. For example, Edelman and Bülthoff (1992) found that in object recognition, generalisation to novel views from a single trained view falls off with increasing angle of rotation. Hill, Schyns, and Akamatsu (1997) also demonstrated that when subjects learned one pose but tested with different ones, generalisation from the learned front pose was progressively worsened as the angle of rotation increased (see also Troje & Bülthoff, 1996; Wallraven, Schwaninger, Schuhmacher, & Bülthoff, 2002). These studies suggest a viewer-centred encoding that depends on a particular vantage point of the observer relative to the pose of a face.

According to view-based theories, encoding several views of an object or face is necessary for pose-invariant recognition. Psychophysical research has found support for these theories by showing that exposures to multiple views of an object or face can facilitate viewpoint or pose-invariant recognition (Edelman & Bülthoff, 1992; Hill et al., 1997; Wallraven et al., 2002; Watson, Johnston, Hill, & Troje, 2005).

Learning to recognise a face in various expressions may require similar exposures to these expressions. Even for familiar faces where recognition is typically expression invariant, unusual expressions can still slow down or hamper recognition performance (Hay, Young, & Ellis, 1991). This suggests that expression-invariant recognition may require certain level of exposures to all basic expressions.

Assuming that learning a face involves encoding both multiple views and expressions, how does the brain integrate the

* Corresponding author. Fax: +86 10 6487 2070.
E-mail addresses: chenwf@psych.ac.cn (W. Chen), c.h.liu@hull.ac.uk (C.H. Liu).
[1] Fax: +44 1482 465599.

0042-6989/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2008.11.003

information from these two types of variation? We should note that this question is different from the literature on the relationship between identity and expression processing, which has centred on the issue of whether identity and expression are mediated by a single or a dual route in the brain. The dual-route theory proposes separate modules for identity and expression processing, drawing evidence from brain-damaged patients whose ability to process one type of information is impaired but the ability to process the other is intact (Bruce & Young, 1986). The single-route hypothesis, on the other hand, cites evidence that recognition of identity can be influenced by recognition of facial expression or vice versa (e.g., Ganel & Goshen-Gottstein, 2004). Over the years since the Bruce and Young's (1986) influential model, the main goal of many studies has been to resolve this debate (Martínez, 2003). However, insights from this line of research have no direct relevance to the chief concern of this study. Here we are mainly interested in how pose and expression-invariant identity recognition is achieved by the visual system. As suggested in the literature, agnosic patients suffering from impaired ability for distinguishing facial expressions can nevertheless have intact ability to recognise faces with various facial expressions (Kurucz & Feldmar, 1979). This shows that expression-invariant recognition of identity is separable from classification of facial expressions although both derive information from non-rigid motion or deformations of the face shape. The main focus of this paper is how these image variations are handled in identity recognition. We aim to address the following question: If several pose and facial expressions need to be learned, does the visual system have to be exposed to each expression in different views? The question may be conceptualised as a matter of transfer between pose and expression training. For instance, if a face has been observed from several poses rather than a single pose, can it be recognised more effectively when it is later seen with a different expression? If the answer is yes, it would suggest that pose training can transfer to a new facial expression. The same question can be asked about the transfer from expression training to a new pose. The purpose of this study is to examine whether training in one type of transformation can be transferred to another.

Experiment 1 examined whether seeing several poses of a face assists recognising the face in a new expression, whereas Experiment 2 examined whether seeing several facial expressions of a face facilitates recognising the face in a new pose. Both experiments employed a sequential matching paradigm where the task was to judge whether a pair of faces presented one after the other were of the same person.

2. Experiment 1

To examine the effect of pose training on matching facial identities with different expressions, we compared performance for conditions where the face at learning was either shown in multiple poses or a single pose. The test face in each trial was either shown with the same or a different expression from the learn face.

2.1. Method

2.1.1. Participants

Twenty undergraduate students from Chinese Agricultural University (mean age 22.8 years, $SD = 1.5$) participated in this experiment. All had normal or correct-to-normal vision.

2.1.2. Materials

The face database was obtained from Binghamton University. It contained 100 3D faces and texture maps without facial hair or spectacles. More details about this database can be found in Yin, Wei, Sun, Wang, and Rosato (2006). We used all the 51 Caucasian and 24 Asian models in the database. Nine additional models were used in the practice session. Each face model was rendered against a black background in seven poses ranging from the full frontal (0°) to six left and right poses (±16°, ±35°, and ±60°). Each pose had seven facial expressions (happiness, sadness, disgust, surprise, anger, fear, and neutral). The rendered faces were saved as grey-level bitmap images. An example face in these variations is shown in Fig. 1. To minimise the low-level image cues for the task, the luminance and root–mean–square contrast of the images were scaled to the grand means. The learn face and the test face were also presented in different sizes, with half of these sized 512 by 512 pixels, whereas the other half 384 by 384 pixels.

2.1.3. Design

We employed a within-participant design. Because our stimuli contained Caucasian faces that could be processed differently by participants of a different race (see, for example, Rhodes, Hayward, & Winkler, 2006), we also included face race as a factor. The independent variables were thus face race (own-race vs. other-race), pose training (multiple pose, single pose, and baseline), and expression change (same vs. different).

2.1.4. Procedure

The experiment was run in two blocks. The pair of faces had the same neutral expression in one block but different expressions in another. Each block consisted of six practice trials and 100 experimental trials. The order of the two blocks was counterbalanced.

Each matching trial consisted of a learn face and a test face presented one after the other in the centre of the screen (see Fig. 2). It began with a 500 ms central fixation cross and a 500 ms blank screen. A learn face was then presented for 3 s. The test face appeared after a 500 ms blank screen. Participants were instructed to judge whether the face images presented at learning and test were of the same person. They were told to give their answer as quickly and accurately as possible by pressing one of the two keys labelled 'yes' or 'no'. The test face remained on screen until the participant responded.

The learn face either consisted of a single or multiple poses of the same person, which always had a neutral expression. In the multiple-pose condition, the six left and right poses were shown successively at 500 ms per pose in the centre of the screen. The pose order was shuffled such that no adjacent poses would be shown consecutively. In the single-pose conditions, the learn face was shown in one of the six side poses. Each pose was assigned randomly with equal frequency. In the baseline condition, the learn face was shown in the full frontal pose.

Each participant completed two blocks of trials, one for the same and another for the different expression. The test face was always a single image with the frontal pose. In the same-expression block, the test face was shown in the same neutral expression as the learn face. In the different-expression block, the test face was shown with an emotional expression. Half of the test faces were the same as the learn face (targets), and the remaining half were different from the learn face (distractors).

2.2. Results

We calculated d' scores for each participant based on the hit and false alarm rates. D' is a parametric measure of sensitivity that indicates how well a participant discriminates targets from distractors. To demonstrate how individual participants performed in this experiment, results from an example participant are presented in Table 1.

The mean d' results across all participants are shown in Fig. 3. A three-way repeated-measures analysis of variance (ANOVA) revealed a significant main effects of pose training, $F(2,38) = 5.65$,