Vision Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Vision Research

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

Expression-dependent susceptibility to face distortions in processing of facial expressions of emotion

Kun Guo[*], Yoshi Soornack, Rebecca Settle

School of Psychology, University of Lincoln, UK

ARTICLE INFO

Number of Reviews = 1

Keywords:
Facial expression
Image resolution
Image blur
Expression categorization
Expression intensity
Gaze behaviour

ABSTRACT

Our capability of recognizing facial expressions of emotion under different viewing conditions implies the existence of an invariant expression representation. As natural visual signals are often distorted and our perceptual strategy changes with external noise level, it is essential to understand how expression perception is susceptible to face distortion and whether the same facial cues are used to process high- and low-quality face images. We systematically manipulated face image resolution (experiment 1) and blur (experiment 2), and measured participants' expression categorization accuracy, perceived expression intensity and associated gaze patterns. Our analysis revealed a reasonable tolerance to face distortion in expression perception. Reducing image resolution up to 48 × 64 pixels or increasing image blur up to 15 cycles/image had little impact on expression assessment and associated gaze behaviour. Further distortion led to decreased expression categorization accuracy and intensity rating, increased reaction time and fixation duration, and stronger central fixation bias which was not driven by distortion-induced changes in local image saliency. Interestingly, the observed distortion effects were expression-dependent with less deterioration impact on happy and surprise expressions, suggesting this distortion-invariant facial expression perception might be achieved through the categorical model involving a non-linear configural combination of local facial features.

1. Introduction

The ability of perceiving and interpreting other people's facial expressions of emotion plays a crucial role in our social interactions, and we are reasonably good at recognizing common facial expressions that represent our typical emotional states and are associated with distinctive pattern of facial muscle movements, such as happy, sad, fear, anger, disgust and surprise (Ekman & Friesen, 1976; Ekman & Rosenberg, 2005), even when these expressive faces appear under very different viewing conditions. For instance, varying face image size to mimic viewing distance in typical social interactions (ranging from arm's length to 5 m; Guo, 2013) or changing face image viewpoint from full frontal view to mid-profile or profile view (Guo & Shaw, 2015; Matsumoto & Hwang, 2011) has little impact on our categorization accuracy of common facial expressions, suggesting the existence of an invariant facial expression representation in our visual system (within given limits) that would be useful for efficient face perception and advantageous to our social interactions.

In addition to view distance and viewpoint, the quality of face image is another common variable we often experience in face perception. Broadly, our visual inputs are not always clean and free of distortions. For instance, we often need to select, extract and process visual information from a noisy environment (e.g., due to rain, snow, fog). Ageing will also decrease optical quality and certain treatments such as refractive surgery will induce significant optical aberrations. Furthermore, the digitized images and videos we view daily (via TV, computer screen, mobile phone, etc.) are subject to a variety of distortions during acquisition, compression, storage, transmission and reproduction (Sheikh, Bovik, & de Veciana, 2005). Any of these causes can result in a degradation of visual quality, typically represented as blurred (e.g., out-of-focus, foggy or raining weather) and low-resolution visual inputs (e.g., images taken by CCTV and mobile phone, or due to compression, denoising and data transmission errors). This raises the question as whether and how we adjust our perceptual processing strategy to compensate degraded visual inputs and maintain a relatively invariant facial expression categorization.

Considering that human visual system has evolved and/or learned over time to process visual signals embedded in natural distortions, it is reasonable to assume that we should have developed certain tolerance to degradation in image quality (Röhrbein, Goddard, Schneider, James, & Guo, 2015). Indeed, psychophysical and computational studies have demonstrated that we could essentially classify natural scenes or

---

[*] Corresponding author at: School of Psychology, University of Lincoln, Lincoln LN6 7TS, UK.
E-mail address: kguo@lincoln.ac.uk (K. Guo).

https://doi.org/10.1016/j.visres.2018.02.001
Received 6 October 2017; Received in revised form 2 February 2018; Accepted 4 February 2018

Please cite this article as: Guo, K., Vision Research (2018), https://doi.org/10.1016/j.visres.2018.02.001

understand scene gist in low-resolution (up to 16 × 16 pixels depending on image complexity) or blurred images (e.g., Castelhano & Henderson, 2008; Torralba, 2009; Watson & Ahumada, 2011). For expressive face images, recent behavioural studies have shown that our expression recognition accuracy remained quite consistent until the image resolution was reduced to around 15 × 10 pixels, in which almost no useful local facial information was left for visual analysis (Du & Martinez, 2011; Johnston, McCabe, & Schall, 2003). However, it is unclear to what extent this 'invariant' facial expression recognition is affected by different types of image distortions, such as image blur. As different facial expressions use different diagnostic spatial frequency bands to transmit expressive facial cues (Smith & Schyns, 2009), image blur may have different impact on expression recognition performance as image resolution. It is also unclear whether the perceived expression intensity, an integral part of expression perception which could lead to changes in behavioural emotional responses (e.g., we may respond differently to the perceived angry expression in low- *vs* high-intensity), is also invariant across the change of face image qualities (e.g., the visibility of facial musculature patterns and local facial features may vary according to image quality, and subsequently affect the perceived intensity for a given expression), and whether the same diagnostic visual cues in low- and high-quality face images are used for expression perception.

Previous eye-tracking research on the effect of image distortion on scene-viewing gaze allocation has generated inconsistent findings. While some studies have suggested that we might use the same diagnostic visual cues in low- and high-resolution scenes (e.g., the location of fixations on low-resolution images tended to be similar to and predictive of fixations on high-resolution images; Judd, Durand, & Torralba, 2011), other studies have observed that viewing of noisy images (e.g., applying masking, low- or high-pass spatial frequency filters to different image regions) was associated with shorter saccade distances, longer fixation durations and stronger central fixation bias (Loschky & McConkie, 2002; Nuthmann, 2013; Röhrbein et al., 2015; van Diepen & d'Ydewalle, 2003), indicating scene-viewing gaze allocation may change with image distortion.

When viewing expressive faces, we tend to scan all key internal facial features (i.e. eyes, nose, and mouth) to extract and then integrate emotional featural cues in order to reliably categorize facial expressions (Guo, 2012, 2013), but look more often at local facial regions that are most characteristic for each facial expression, such as mouth in happy faces and eyes in angry faces (Eisenbarth & Alpers, 2011; Jack, Blais, Scheepers, Schyns, & Caldara, 2009; Schurgin et al., 2014). If this pattern of gaze distribution is tightly coupled with expression categorization performance, then the image quality-invariant expression perception would suggest consistent gaze pattern across different face image distortions. If, on the other hand, this pattern of gaze distribution could be dissociated with expression categorization performance, then face distortion might significantly affect our face-viewing gaze allocation.

In separate eye-tracking experiments with a self-paced expression categorization task, we systematically investigated our perceptual sensitivity (categorization accuracy, perceived expression intensity, and reaction time) to face image resolution and blur. We aimed to examine the extent to which (1) expression perception is invariant across different degrees and types of face distortions, (2) there is a tight correspondence between specific gaze allocation and expression recognition performance.

2. Experiment 1: Effect of image resolution on facial expression categorization

2.1. Methods

Twenty-six undergraduate students (12 male, 14 female), age ranging from 18 to 23 years old with the mean of 19.85 ± 0.29 (Mean ± SEM), volunteered to participate in this study. This sample size is compatible with previous studies in this area (e.g., Guo, 2012, 2013; Jack et al., 2009). All participants had normal or corrected-to-normal visual acuity. The Ethical Committee in School of Psychology, University of Lincoln approved this study. Written informed consent was obtained from each participant, and all procedures complied with the British Psychological Society Code of Ethics and Conduct, and were in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

The general experimental setup and testing procedure has been described in our previous publications (e.g., Green & Guo, 2018; Guo & Shaw, 2015). Briefly, digitized grey-scale face images in full frontal view were presented through a ViSaGe graphics system (Cambridge Research Systems, UK) and displayed on a non-interlaced gamma-corrected colour monitor (30 cd/m^2 background luminance, 100 Hz frame rate, Mitsubishi Diamond Pro 2070SB) with the resolution of 1024 × 768 pixels. At a viewing distance of 57 cm, the monitor subtended a visual angle of 40 × 30°.

Western Caucasian face images, consisting of four female and four male models, were selected from the Karolinska Directed Emotional Faces CD ROM (Lundqvist, Flykt, & Öhman, 1998). Each of these young adult models posed six common facial expressions (happy, sad, fearful, angry, disgusted, and surprised). Although they may have real-world limitations, and categorization performance for some expressions could be subject to culture influence, these well-controlled face images were chosen for their comparability and universality in transmitting facial expression signals, at least for our age-matched observer group (Western Caucasian young adults). The faces were processed in Adobe Photoshop to remove external facial features (e.g., hair), to ensure a homogenous grey background, and then to downsize to 384 × 512 pixels (referred as resolution 1). For each of these 'resolution 1' face images, three subsequent faces were constructed by further downsizing to 48 × 64 pixels (resolution 1/8), 24 × 32 pixels (resolution 1/16), 12 × 16 pixels (resolution 1/32). To provide a constant presentation size for all face images, the three downsized faces were scaled back to 384 × 512 pixels (14 × 19°) using bilinear interpolation, which preserves most of the spatial frequency components. As a result, 192 expressive face images were generated for the testing session (4 face resolutions × 6 expressions × 8 models, see Fig. 1 for examples). These images were gamma corrected and displayed once in a random order during the testing to minimise the potential practice or carryover effects (e.g., exposing to the same face identity displaying the same expression in consecutive trials).

All of our participants were aware of universal facial expressions. Before the testing, they were shown a PowerPoint presentation containing one male and one female model posing happiness, sadness, fear, anger, disgust, and surprise (sampled from Pictures of Facial Affect), and were asked to label each facial expression as carefully as possible without time constraint. All of them could recognize these facial expressions or agree with the classification proposed by Ekman and Friesen (1976).

A self-paced task was used to mimic natural viewing condition. During the experiments the participants sat in a chair with their head restrained by a chin-rest, and viewed the display binocularly. To calibrate eye movement signals, a small red fixation point (FP, 0.3° diameter, 15 cd/m^2 luminance) was displayed randomly at one of 9 positions (3 × 3 matrix) across the monitor. The distance between adjacent FP positions was 10°. The participant was instructed to follow the FP and maintain fixation for 1 s. After the calibration procedure, the participant pressed the response box to initiate a trial. The trial was started with an FP displayed 10° left or right to the screen centre to minimize central fixation bias (Tatler, 2007). If the participant maintained fixation for 1 s, the FP disappeared and a face image was presented at the centre of the monitor. During the self-paced, free-viewing presentation, the participant was instructed to "categorize this facial expression as accurately and as quickly as possible", and to respond by pressing a