Vision Research 150 (2018) 29–37

Contents lists available at ScienceDirect

Vision Research

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

Orientation information in encoding facial expressions

Deyue Yu[a,*], Andrea Chai[b], Susana T.L. Chung[b]

[a] College of Optometry, The Ohio State University, Columbus, OH, United States
[b] School of Optometry, University of California, Berkeley, CA, United States

ARTICLE INFO

Number of reviewers : 2

Keywords:
Spatial vision
Psychophysics
Facial expression recognition

ABSTRACT

Previous research showed that we use different regions of a face to categorize different facial expressions, e.g. mouth region for identifying happy faces; eyebrows, eyes and upper part of nose for identifying angry faces. These findings imply that the spatial information along or close to the horizontal orientation might be more useful than others for facial expression recognition. In this study, we examined how the performance for recognizing facial expression depends on the spatial information along different orientations, and whether the pixel-level differences in the face images could account for subjects' performance. Four facial expressions—angry, fearful, happy and sad—were tested. An orientation filter (bandwidth = 23°) was applied to restrict information within the face images, with the center of the filter ranged from 0° (horizontal) to 150° in steps of 30°. Accuracy for recognizing facial expression was measured for an unfiltered and the six filtered conditions. For all four facial expressions, recognition performance (normalized d') was virtually identical for filter orientations of −30°, horizontal and 30°, and declined systematically as the filter orientation approached vertical. The information contained in mouth and eye regions is a significant predictor for subject's response (based on the confusion patterns). We conclude that young adults with normal vision categorizes facial expression most effectively based on the spatial information around the horizontal orientation which captures primary changes of facial features across expressions. Across all spatial orientations, the information contained in mouth and eye regions contributes significantly to facial expression categorization.

1. Introduction

Recognizing facial expressions is an important skill in social interactions. Many previous studies have focused on evaluating the role of spatial frequencies in facial expression recognition. It has been shown that subjects perform the task of categorizing facial expressions based on low spatial frequency information contained within the face images, although the detection of an expression and the strength of an expression engages the use of high spatial frequency information (Calder, Young, Keane, & Dean, 2000; Schyns & Oliva, 1999; Vuilleumier, Armony, Driver, & Dolan, 2003). In early visual processing, retinal input such as face image was decomposed not only along the dimension of spatial frequency but also along the dimension of orientation (e.g., (De Valois, Albrecht, & Thorell, 1982; Hubel & Wiesel, 1968). We have also learnt that to precisely categorize facial expressions, subjects tend to use different configurations of facial regions (e.g. mouth region for happy face; eyebrows, eyes and upper part of nose for angry face) (Smith, Cottrell, Gosselin, & Schyns, 2005). Most of the facial elements such as eyebrows, eyes, mouths are more horizontally oriented. When generating facial expressions, there also seems to be more variations in configurations among facial elements oriented horizontally. As shown in a later section (Figs. 6–8), image comparisons between different facial expressions demonstrate that the majority of the configural differences occur near horizontal orientation. These findings indicate that information conveyed by channels along or near horizontal orientation might be more important than the others for facial expression recognition.

Several studies explored how the orientation of spatial information could affect face identification (e.g., Dakin & Watt, 2009; Yu & Chung, 2011). By evaluating face identification using images filtered along various bands of orientation, these studies showed that subjects performed best when viewing images containing information close to horizontal orientation, with performance declining gradually as the orientation of the reserved information approached vertical. Goffaux and Dakin (2010) further examined the impact of horizontally orientated facial information on several key behavioral signatures of face perception: inversion effect, identity after-effect, matching across viewpoints, and interactive processing of parts. They found that preferential processing of information around the horizontal orientation provides a significant account of the behavioral measures of face

* Corresponding author.
E-mail address: deyueyu@gmail.com (D. Yu).

https://doi.org/10.1016/j.visres.2018.07.001
Received 23 November 2017; Received in revised form 25 June 2018; Accepted 12 July 2018
0042-6989/

processing. While the invariant aspects of faces encode face identity, the changeable aspects of faces construct emotional expressions (Haxby, Hoffman, & Gobbini, 2000). It has been suggested that separate functional and neural pathways are involved in the perception of invariant aspects of faces and of changeable aspects of faces (Bruce & Young, 1986; Hasselmo, Rolls, & Baylis, 1989; Haxby et al., 2000; Winston, Henson, Fine-Goulden, & Dolan, 2004), implying that identifying faces and categorizing facial expressions could depend on different input information. On the other hand, both face identification and facial expression categorization have been shown to rely on the configural information of facial components (Calder et al., 2000; Leder, Candrian, Huber, & Bruce, 2001). Therefore, it remains unclear whether the spatial information most crucial for categorizing facial expressions is the same as that for recognizing face identities.

A recent study utilized orientation bubbles to reveal the diagnostic information for facial expressions and found a strong link between the horizontal information and the successful categorization of several facial expressions (anger, disgust, fear, happy and sad) but not for the surprise expression (Duncan et al., 2017). These authors further showed that individual differences in the reliance of horizontal information were best predicted by the utilization of eye region alone. However, facial regions other than the eyes have been shown to be important for expression categorization. In fact, Smith et al (2005) showed that the facial regions diagnostic of a certain emotion expression are different for different expressions and share very little overlapping in their locations on a face image. Also, by examining only happy and sad expressions, Huynh and Balas (2014) found that the magnitude of the preference of horizontal orientation (compared to vertical) can be modulated by factors such as mouth openness.

In this study, we systematically evaluated the dependency of facial expression categorization on the orientation of spatial information. Specifically, we examined how the performance for recognizing facial expression depends on information restricted to different orientation bands. We asked whether categorizing facial expressions shows a similar orientation dependency on spatial information as that for recognizing face identities, i.e. primarily the horizontal structures. In addition, we examined the confusion patterns among different facial expressions for different filter orientations, and investigated how local facial regions (differences between facial expressions at the pixel level) may contribute to the categorization performance.

2. Methods

2.1. Subjects

Fifteen subjects (eight females and seven males) with normal or corrected-to-normal vision, aged between 18 and 39 years, participated in the study. All subjects were naïve to the purpose of the experiment, and performed the task binocularly. The research was conducted in accordance with the Declaration of Helsinki. Prior to the commencement of data collection, every subject signed a consent form approved by the Institutional Review Board at the University of California, Berkeley.

2.2. Apparatus and stimuli

We used custom-written software written in MATLAB (version 7.7.0) and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) to control the experiment using a Macintosh computer (MacBook 5.1). Stimuli were presented on a gamma-corrected SONY color graphic display (model: Multiscan E540), at a pixel resolution of 1280 × 1024 (dimensions: 39.3 cm × 29.4 cm) and a refresh rate of 75 Hz.

Four facial expressions were tested: angry, fearful, happy and sad. We selected stimuli from the NimStim Set of Facial Expressions, a standardized database of naturally posed photographs of professional actors (Tottenham et al., 2009). As shown by Huynh and Balas (2014), the openness of the mouth can influence the emotion-dependent reliance on horizontally orientated face information. To examine the effect of filter orientation without the possible interfering effect of mouth openness, only closed-mouth versions were used in the study. To ensure none of the subjects viewed the same image more than once, we generated more test faces by morphing (Abrosoft FantaMorph 4 Deluxe) between two persons (of the same gender) with the same facial expression (a total of 118 source images were used). There were a total of 140 different faces (morphed and original) obtained for each facial expression, with 55 female faces and 85 male faces. To create a morphing image, about 100 key dots were placed on the major elements (such as eyes, eyebrows, nose, mouth, the outline of the face, and creases induced by facial expressions) of both source images. Each key point on one face image was automatically matched to its corresponding key point on the other face. The two source face images were then linearly interpolated by the software to produce a morphed image. Only one morphing level, 50%, was used so that the facial features from both faces were equally presented. Additionally, for each image (morphed and original), two reference points were defined, one at the center of the mouth and the other at the midpoint between the eyes. Rotation was then made to each image until the two reference points fell on a vertical line. The mean distance between the two reference points was 153 pixels for happy, 167 pixels for sad, 162 pixels for angry, and 167 pixels for fearful.

An orientation filter (wrapped Gaussian distributions with a bandwidth $\sigma = 23^\circ$) was applied to restrict information contained in the stimuli, with the center of the filter ranged from 0° (horizontal) to 150° in steps of 30°, as in Yu and Chung (2011). For each filtered condition, information within the filter orientation ± the bandwidth was retained, and the rest was filtered out. All face images were converted to gray scale and cropped to an oval shape (minor and major diameters are 273 and 405 pixels). Across all conditions, images were normalized to equate the root mean square (RMS) contrast (0.12) and luminance (0.5). Stimuli were presented on a gray background (29 cd/m^2). Subjects were tested binocularly. At our viewing distance of 40 cm, the angular subtense of the images was 8° horizontally and 11.9° vertically. Fig. 1 shows examples of the four facial expressions in the unfiltered and the six filtered conditions. Accuracy for recognizing facial expressions filtered with each of these filters, as well as for the unfiltered condition, was measured.

2.3. Procedures

There were a total of 28 conditions (four facial expressions × seven filter orientations). Twenty trials per filter orientation were tested for each facial expression and each subject. For each subject, trials were divided into four blocks with 140 trials per block (testing conditions were completely randomized within each block). No subject viewed the same image more than once. Prior to testing, each subject completed a practice block using a different set of face images to familiarize themselves with the task.

Before each trial, a white fixation dot was presented at the center of the display. Subjects were instructed to press a mouse button to initiate a trial. Each face image was presented for 53 ms, which was selected based on pilot data to avoid any ceiling or floor effect in performance. Immediately after the stimulus disappeared, a white-noise post-mask was presented for 500 ms, followed by a response screen displaying four choices in words—angry, fearful, happy and sad. Using the mouse, subjects selected the response for each trial. Fig. 2 illustrates a schematic diagram of the experimental paradigm.

3. Results

The proportion correct of recognition, averaged across the 15 subjects, was plotted as a function of the orientation of the spatial filter for each facial expression in Fig. 3. Given that the task was a four