ARTICLE IN PRESS

Vision Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Vision Research

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

Capacity limitations to extract the mean emotion from multiple facial expressions depend on emotion variance

Luyan Ji*, Gilles Pourtois

Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium

ARTICLE INFO

Number of Reviews = 2

Keywords:
Ensemble representation
Facial expressions
Processing capacity limitations
Set size
Amplifying effect
Sampling

ABSTRACT

We examined the processing capacity and the role of emotion variance in ensemble representation for multiple facial expressions shown concurrently. A standard set size manipulation was used, whereby the sets consisted of 4, 8, or 16 morphed faces each uniquely varying along a happy-angry continuum (Experiment 1) or a neutral-happy/angry continuum (Experiments 2 & 3). Across the three experiments, we reduced the amount of emotion variance in the sets to explore the boundaries of this process. Participants judged the perceived average emotion from each set on a continuous scale. We computed and compared objective and subjective difference scores, using the morph units and post-experiment ratings, respectively. Results of the subjective scores were more consistent than the objective ones across the first two experiments where the variance was relatively large, and revealed each time that increasing set size led to a poorer averaging ability, suggesting capacity limitations in establishing ensemble representations for multiple facial expressions. However, when the emotion variance in the sets was reduced in Experiment 3, both subjective and objective scores remained unaffected by set size, suggesting that the emotion averaging process was unlimited in these conditions. Collectively, these results suggest that extracting mean emotion from a set composed of multiple faces depends on both structural (attentional) and stimulus-related effects.

1. Introduction

For the last ten years, evidence has accumulated showing that human observers are able to rapidly process multiple emotional faces shown concurrently and extract the average emotion from them (e.g., Elias, Dyer, & Sweeny, 2017; Haberman & Whitney, 2007, 2009; Ji, Rossi, & Pourtois, 2018). The representation which summarizes multiple features or items into an ensemble is referred to as ensemble representation (Alvarez, 2011; Whitney & Leib, 2018), and is thought to allow outlier detection in visual search (Cavanagh, 2001), as well as minimize the impression of being exposed to a visual world that would be too rich and complex to handle (Cohen, Dennett, & Kanwisher, 2016; Rensink, O'Regan, & Clark, 1997).

Like averaging low-level features or stimuli, for example orientation (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001) and size (Ariely, 2001; Chong & Treisman, 2005), the ability of deriving the affective gist from multiple facial expressions has also been shown to be very robust and flexible across different tasks and contexts, occurring implicitly (Haberman & Whitney, 2007), and even on sets containing as many as 24 individual faces shown simultaneously for only 100 ms (Yang, Yoon, Chong, & Oh, 2013). In addition, even when the accuracy of individual representations is very low (e.g., at chance level) because of limited attentional resources, ensemble representation remains surprisingly precise (Fischer & Whitney, 2011; Haberman & Whitney, 2009, 2011; Li et al., 2016).

On the other hand, the underlying perceptual mechanism responsible for creating ensemble representation for higher-level information (such as facial expressions) is still largely unclear and under debate in the existing literature. An open question remaining pertains to knowing whether ensemble representation could help overcome or bypass limitations in visual processing (Alvarez, 2011; Attarha, Moore, & Vecera 2014; but cf. Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013; Chong & Treisman, 2005; Cohen et al., 2016; Ji, Chen, Loeys, Pourtois, 2018).

One way to assess attention bottlenecks in visual processing is using a classical set-size (i.e., the number of items in the set) manipulation (Theeuwes, 1992; Treisman & Gelade, 1980; Wolfe, 2007). This manipulation has been widely used in visual search studies in the past. For example, searching for a negative (angry) face surrounded by neutral faces used as distractor was found to be less impaired by increasing set sizes and thus more efficient, compared to a control condition where a positive face had to be searched in the set (Horstmann, 2007; Öhman,

* Corresponding author at: Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent 9000, Belgium.
E-mail address: Luyan.Ji@Ugent.be (L. Ji).

https://doi.org/10.1016/j.visres.2018.03.007
Received 6 October 2017; Received in revised form 23 January 2018; Accepted 11 March 2018

Lundqvist, & Esteves, 2001). Based on capacity models for divided attention (Broadbent, 1958; Kahneman, 1973), if visual processing is capacity unlimited, then each stimulus is analyzed independently, so that the quality of perception does not vary (i.e., decrease) with increasing set sizes. In comparison, if perceptual capacity is limited,[1] then there is by definition a limit imposed on the amount of information processed at a given time such that interference (competition) between stimuli occurs, particularly so when the number of stimuli in the set increases.

Different from the earlier psychophysical studies which focused on exploring set size effect in the processing and detection of a single target, studies on ensemble representation consider all stimuli in the set as target elements that all participate in principle to shape visual processing and eventually determine emotion perception. Using set-size manipulations, Haberman and Whitney (2007, 2009) previously found that the averaging performance was not influenced by increasing set sizes, which provided support for a capacity unlimited process, and was consistent with findings on averaging low-level features or stimuli (e.g., mean size, Ariely, 2001; Chong & Treisman, 2005). Similar to these previous psychophysical studies on mean size representation (Ariely, 2001), Haberman and Whitney (2007, 2009) used a uniform distribution of emotional intensities composed of four unique morph units, and notably, they selected only one single identity. Therefore, the sets used in their study usually remained relatively homogeneous or regular with only four different expressions, no matter whether the set size was 4 or 16. However, these homogenous sets would presumably ease the sampling strategies. As a matter of fact, in these conditions, sampling only one or two items might explain behavioral performance and the resistance to set size manipulations, as demonstrated by simulation methods (Myczek & Simons, 2008).

In a recent study (Ji, Chen, et al., 2018), we found that the perceptual capacity of establishing mean representation for mixed full-blown angry and happy facial expressions was limited, using the extended simultaneous-sequential paradigm (Scharff, Palmer, & Moore, 2011). However, it might be challenging and also uncommon to average multiple facial expressions that convey distinct, and even opposite, emotion categories (i.e., happiness vs. anger), as the variance in the set is necessarily high in these conditions. Further, it has been shown previously with low-level attributes such as size or orientation that the averaging turned out to be easier and more accurate when the variance in the set was reduced (e.g., Solomon, Morgan, & Chubb, 2011). To overcome this problem, in the current study, we used a standard morphing technique meant to reduce the variance of facial expressions presented within the set, as well as to better control their actual emotion intensity values. Moreover, we also manipulated this factor across different experiments to examine if it reliably influenced the averaging process. On the other hand, in order to reduce the regularity in the set and thus create a situation where a subsampling strategy would be inadequate to perform the averaging task, we decided to use different stimuli in the set invariably, namely having different emotional values each time, as was done previously in the case of mean size perception (Marchant, Simons, & de Fockert, 2013; Utochkin & Tiurina, 2014). However, for emotional facial expressions that have a more limited range than low-level properties, a caveat is that for larger set sizes, they are still rather homogeneous as the different stimuli composing the set are necessarily similar. As a compromise, in the current study, we employed a uniform distribution of four unique morph units, regardless of the varying set size (from 4 to 16), similarly to Haberman and Whitney (2007, 2009), but unlike them, we selected 16 different face identities, to increase heterogeneity in the set. In addition, unlike Haberman and Whitney (2007, 2009), we also collected from the same participants emotion ratings for all the individual (unmorphed) faces used in the main experiment in order to assess whether the objective (i.e., actual morph unit) or subjective (i.e., valence intensity rating) value best accounted for the averaging performance during the task (see Methods for details). This choice was motivated by the results of our previous study (Ji, Chen, et al., 2018) where we found that the subjective emotion perception of faces was a reliable predictor of performance during the main averaging task since it took into account the subject-specific perception of the emotional faces used as stimuli that can vary considerably across participants (unlike fixed morph units).

All in all, the current study therefore aimed at exploring the (attention) boundaries for extracting the mean emotion from a set composed of multiple facial expressions and how the emotion variance across them could modulate the processing capacity, using a standard set size manipulation and well controlled face stimuli (by means of a morphing procedure). To this aim, three different experiments were performed. Across them, participants judged the perceived average emotion from each face set on a continuous scale (similarly to Ji, Chen, et al., 2018). The face set consisted of 4, 8 or 16 faces, and was presented for 500 ms. In Experiment 1, we used morphed faces extracted from a continuum going from anger to happiness, hence providing a between-emotion categories manipulation. In Experiments 2 and 3, we used within-emotion continua (either from neutral to happy or from neutral to angry) in separate blocks, to decrease the inter-item (face) variance in the sets in terms of emotional expressions. Further, Experiment 3 differed from Experiment 2 in that the distance between the different morph units was smaller (thus the emotion intensity variance within the face set was smaller) in the former compared to the latter experiment. (i) We predicted that the averaging performance should mainly be capacity-limited (see Ji, Chen, et al., 2018), in the sense of being influenced by the set size manipulation: a worse performance was expected for large compared to small set sizes. (ii) In addition, we hypothesized that the averaging performance would improve and be less affected by set size when the inter-item (face) variance (in terms of emotion expressions) decreased. Hence, we surmised modulatory effects of set size and inter-stimulus variance on the ability to extract the mean emotion from a complex set composed of multiple facial expressions.

2. General methods

2.1. Participants

All three experiments included twenty-four participants from Ghent University (Experiment 1: 18–25 years, 17 females; Experiment 2: 18–25 years, 15 females; Experiment 3: 19–28 years, 19 females). The sample size of 24 was determined a priori to be consistent with our previous behavioral study (see Ji, Chen, et al., 2018). The participants gave written informed consent prior to the start of the experiment and were compensated 10 Euro per hour. They reported to be right-handed and have normal or corrected-to-normal vision. The study protocol was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee.

2.2. Stimuli

Sixteen different identities, eight males and eight females, were selected from the NimStim database (Tottenham et al., 2009). Each face identity showed happy, angry, or neutral expression, all with closed mouth. The hair, ears, neck and other external information were cropped. All images were converted to grey scale, and scaled to the same mean luminance and root-mean-square contrast (Bex & Makous, 2002). Each face image subtended a visual angle $4.03^\circ \times 4.28^\circ$, and was presented against a homogenous black background.

Face images were generated by morphing using FantaMorph 5. In

[1] A limited-capacity parallel model is also possible (Palmer, 1990), but in the current study, we did not distinguish between parallel and serial accounts of limited-capacity processing as it goes beyond its scope.