NeuroImage 184 (2019) 813–825

Contents lists available at ScienceDirect

NeuroImage

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

Multimodal evidence on shape and surface information in individual face processing

Dan Nemrodov[a,*], Marlene Behrmann[b,c], Matthias Niemeier[a], Natalia Drobotenko[a], Adrian Nestor[a]

[a] Department of Psychology at Scarborough, University of Toronto, 1265 Military Trail, Toronto, Ontario, M1C1A4, Canada
[b] Department of Psychology at Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213, USA
[c] Center for the Neural Basis of Cognition, University of Pittsburgh, PA 4400 Fifth Ave, Pittsburgh, PA, 15213, USA

ARTICLE INFO

Keywords:
Face recognition
Image reconstruction
fMRI
EEG

ABSTRACT

The significance of shape and surface information for face perception is well established, yet their relative contribution to recognition and their neural underpinnings await clarification. Here, we employ image reconstruction to retrieve, assess and visualize such information using behavioral, electroencephalography and functional magnetic resonance imaging data.

Our results indicate that both shape and surface information can be successfully recovered from each modality but that the latter is better recovered than the former, consistent with its key role for face representations. Further, shape and surface information exhibit similar spatiotemporal profiles, rely on the extraction of specific visual features, such as eye shape or skin tone, and reveal a systematic representational structure, albeit with more cross-modal consistency for shape than surface. More generally, the present work illustrates a novel approach to relating and comparing different modalities in terms of perceptual information content.

Thus, our results help elucidate the representational basis of individual face recognition while, methodologically, they showcase the utility of image reconstruction and clarify its reliance on diagnostic visual information.

1. Introduction

The segregation of shape and surface information defines a fundamental aspect of visual processing and cortical organization (Livingstone and Hubel, 1988; Van Essen and Deyoe, 1995) both in the human (Cant et al., 2008; Lafer-Sousa et al., 2016; Vinberg and Grill-Spector, 2008) and the monkey brain (Conway et al., 2007). Accordingly, this distinction has played a prominent role in accounts of face recognition (Bruce and Young, 1998). Extensive research has documented the importance of both types of information in face perception (Biederman and Kalocsai, 1997; Jiang et al., 2006; O'Toole et al., 1999; Russell et al., 2007; Russell and Sinha, 2007; Vuong et al., 2005), but the relative weight of shape and surface properties has been heavily debated, with either the former (Jiang et al., 2011; Lai et al., 2013) or the latter (Bruce et al., 1991; Bruce and Langton, 1994; Hole et al., 2002; Kaufmann and Schweinberger, 2008; Russell et al., 2006) considered dominant. Arguably, this debate arises from a lack of specificity in identifying the shape and surface features critical for individual face processing (Burton et al., 2015). Thus, the current research aims to uncover the nature of the information involved in individual face processing along with its accompanying neural profile.

To address the challenge above, here, we appeal to neural-based image reconstruction (Shen et al., 2018; Miyawaki et al., 2008; Naselaris et al., 2009; Nishimoto et al., 2011a,b; Thirion et al., 2006), namely, the endeavor of reconstructing the appearance of visual objects from neural activity prompted by their processing. While this endeavor has relied primarily on functional magnetic resonance imaging (fMRI), more recently, additional modalities have been used successfully as well. For instance, facial image reconstruction has been carried out using single-cell recordings (Chang and Tsao, 2017), electroencephalography (EEG) data (Nemrodov et al., 2018) and behavioral data (Chang et al., 2017b; Zhan et al., 2017), in addition to fMRI (Cowen et al., 2014; Lee and Kuhl, 2016; Nestor et al., 2016). Thus, in theory, image reconstruction can provide a powerful platform for investigating shape/surface processing in face individuation via multiple behavioral and neuroimaging modalities. Concretely, image reconstruction can be used to uncover, assess and compare facial shape and surface information recovered from distinct modalities.

* Corresponding author.
E-mail address: dan.nemrodov@utoronto.ca (D. Nemrodov).

https://doi.org/10.1016/j.neuroimage.2018.09.083
Received 23 August 2018; Received in revised form 22 September 2018; Accepted 30 September 2018
Available online 3 October 2018

To this end, we rely on data assessing individual face processing gleaned from behavioral (Nestor et al., 2013), EEG (Nemrodov et al., 2018) and fMRI data (Nestor et al., 2016). Specifically, for each modality, we aim to recover the shape and surface content of a common set of face stimuli as perceived by human observers. In addition, the same procedure is conducted with an image-based theoretical observer (TO) allowing us to compare the informational content of multiple empirical and TO reconstructions.

To achieve these goals, we appeal to an influential approach for analyzing face images into shape and surface properties (Craw and Cameron, 1991; Kramer et al., 2016; Tiddeman et al., 2001; Vetter and Troje, 1995). Specifically, this approach involves marking the positions of a set of fiducial points (e.g., the corners of the eyes or the tip of the nose) that deliver shape information. Then, faces are warped to a standard shape (i.e., a preset configuration of fiducial points) yielding 'shape-free' images that deliver surface information. To be clear, shape derived in this manner encompasses two sources of information: configural information, conceived as metric distances between different face parts (Maurer et al., 2002; Tanaka and Gordon, 2011), and local information associated with the geometric structure of specific face parts such as eye shape or mouth shape (Cabeza and Kato, 2000; Gold et al., 2012; Rakover, 2002). In contrast, surface contains information about the reflectance properties of a face (e.g., hue, specularity, albedo) that also play a role in individual face recognition (Hancock et al., 1996; Russell et al., 2007; Taschereau-Dumouchel et al., 2010) – such information is alternatively referred to as 'texture', 'pigmentation' or 'surface reflectance'.

The appeal to shape-surface decomposition allows us to address a number of related questions. First, can image reconstruction separately recover facial shape and surface information from different modalities and, if so, how well? Second, what is the spatiotemporal profile of shape and surface processing? Third, what specific shape/surface features are recovered through reconstruction? And fourth, do different modalities reveal similar or complementary information about face representations? More generally, the present work evaluates and confirms the ability of a novel methodological paradigm to exploit multimodal evidence in an effort to elucidate the representational content of individual face processing.

In summary, the current work embarks on a comprehensive investigation of facial shape and surface processing by appealing to powerful and innovative image-reconstruction methodology as applied to multimodal data. Accordingly, this work serves a twofold purpose by shedding light on the psychological and neural profile of facial shape/surface processing and by clarifying the informational content responsible for the success of image reconstruction.

2. Materials and methods

2.1. Stimuli

A common subset of 108 stimulus images was identified across three different studies investigating empirical and computational aspects of unfamiliar face recognition (see 2.3 Experimental procedures). Images of 54 individuals displaying neutral and happy facial expressions were selected from three databases: AR (Martinez and Benavente, 1998), FEI (Thomaz and Giraldi, 2010) and Radboud (Langner et al., 2010). All images featured young adult Caucasian males with frontal view, gaze and illumination. The stimuli were selected so that no facial accessories, hair or makeup obscured the internal features of the face and so that all happy expressions displayed an open-mouth smile. These images were: (a) scaled uniformly and aligned with roughly the same position of the eyes and the nose; (b) cropped to eliminate background; (c) normalized with the same mean and root mean square (RMS) contrast values separately for each color channel in CIEL*a*b* color space, and (d) reduced to the same size (95 × 64 pixels). Note that this procedure did not change the aspect ratio of the images though the position of the eyes and the nose was roughly the same across stimuli. Thus, every effort was made to homogenize the stimulus set both in terms of low-level and high-level face properties preventing the potential contribution of such factors to image reconstruction.

2.2. Participants

All participants (age range across studies: 18–34 years; 21 males, 22 females) were Caucasian adults with normal or corrected-to-normal vision and no history of cognitive or neurological disorder. All participants provided informed consent and all experimental procedures were approved by the Research Ethics Board at University of Toronto and/or the Institutional Review Board at Carnegie Mellon University.

2.3. Experimental procedures

Data used for reconstruction purposes were selected from three previous studies as follows.

Behavioral data consisted of similarity ratings with pairs of faces acquired from 22 participants (reported in Nestor et al., 2013, Experiment 1). Briefly, on each trial, participants were presented with two facial identities displaying different emotional expressions, one neutral and one happy, side by side, for 400 ms, and were asked to judge their visual similarity on a 5-point scale. Each participant rated all possible 1431 facial pairs, corresponding to 54 facial identities - for clarity, only a subset of the original data were considered here (i.e., 6 additional facial identities were not used in the EEG study summarized below and, hence, were excluded from further analyses of behavioral data).

EEG data were previously acquired from 13 participants who performed a go/no-go gender categorization task (Nemrodov et al., 2018). On 'no-go' trials, participants viewed the stimuli described above while, on 'go' trials, they were asked to press a designated key in response to the appearance of a female face. Each of the 108 main stimuli was presented for 300 ms and repeated across 64 trials for each participant.

fMRI data were acquired from 8 participants who performed a continuous one-back identity task (Nestor et al., 2016). Briefly, on each trial, participants viewed a stimulus for 900 ms and responded whether the current stimulus displayed the same individual as that presented on the previous trial, irrespective of emotional expression. The experiment used a wide-spaced design (8s trials) and allowed for the repetition of each stimulus for a minimum of 10 trials across five 1-hr sessions for each participant. Again, only a subset of the stimuli used in the original study is considered here to enable direct comparison with data from the other modalities.

To be clear, we note that the neuroimaging studies above (Nemrodov et al., 2018; Nestor et al., 2016) did not separate shape and surface cues for reconstruction purposes nor did they assess the contribution of such cues to visual face representations. Further, the behavioral study above (Nestor et al., 2013) did not target any form of image reconstruction and, thus, it provides a new testing ground for reconstruction endeavors.

2.4. Representational similarity analyses

Our reconstruction procedure fundamentally relies on the structure of representational (dis)similarity matrices (Kriegeskorte et al., 2008) to derive facial image features and to use such features for reconstruction purposes. Hence, the first step of our investigation is to construct such matrices separately for each data type.

Specifically, for each modality and for each participant, a similarity matrix was designed to store pairwise similarity estimates across 54 facial identities. In the case of behavioral data, these estimates were readily available in the form of similarity ratings. In the case of EEG and fMRI data such estimates were derived through one-against-one pattern classification of different identities, separately for each expression, using linear support vector machines (SVM). Briefly, pairwise classification was applied across EEG spatiotemporal patterns recorded across at 12