Schwaninger, A., Lobmaier, J. S., & Collishaw, S. M. (2002). *Lecture Notes in Computer Science, 2525,* 643-650.

Role of Featural and Configural Information in Familiar and Unfamiliar Face Recognition

Adrian Schwaninger[1), 2)]*, Janek S. Lobmaier[2)], and Stephan M. Collishaw[3)]

[1)] Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2)] Department of Psychology, University of Zürich, Switzerland
[3)] School of Cognitive and Computing Sciences, University of Sussex, UK

Abstract. Using psychophysics we investigated to what extent human face recognition relies on local information in parts (featural information) and on their spatial relations (configural information). This is particularly relevant for biologically motivated computer vision since recent approaches have started considering such featural information. In Experiment 1 we showed that previously learnt faces could be recognized by human subjects when they were scrambled into constituent parts. This result clearly indicates a role of featural information. Then we determined the blur level that made the scrambled part versions impossible to recognize. This blur level was applied to whole faces in order to create configural versions that by definition do not contain featural information. We showed that configural versions of previously learnt faces could be recognized reliably. In Experiment 2 we replicated these results for familiar face recognition. Both Experiments provide evidence in favor of the view that recognition of familiar and unfamiliar faces relies on featural and configural information. Furthermore, the balance between the two does not differ for familiar and unfamiliar faces. We propose an integrative model of familiar and unfamiliar face recognition and discuss implications for biologically motivated computer vision algorithms for face recognition.

Introduction

Different object classes can often be distinguished using relatively distinctive features like color, texture or global shape. In contrast, face recognition entails discriminating different exemplars from a quite homogeneous and complex stimulus category. Several authors have suggested that such expert face processing is holistic, i.e. faces are meant to be encoded and recognized as whole templates without representing parts explicitly [4,5,6]. In computer vision many face recognition algorithms process the whole face without explicitly processing facial parts. Some of these algorithms have been thought of being particularly useful to understand human face recognition and were cited in studies that claimed faces to be the example for exclusive holistic processing (e.g. [7,8] cited in [9], or the computation models cited in [6], p. 496).

* AS was supported by a grant from the European Commission (IST Programme).

In contrast to holistic algorithms like principal components analysis or vector quantization, recent computer vision approaches have started using local part-based or fragment-based information in faces [1,2,3]. Since human observers can readily tell the parts of a face such algorithms bear a certain intuitive appeal. Moreover, potential advantages of such approaches are greater robustness against partial occlusion and less susceptibility to viewpoint changes.

In the present study we used psychophysics to investigate whether human observers only process faces holistically, or whether they encode and store the local information in facial parts (featural information) as well as their spatial relationship (configural information). In contrast to previous studies, we employed a method that did not alter configural or featural information, but eliminated either the one or the other. Previous studies have often attempted to directly alter the facial features or their spatial positions. However, the effects of such manipulations are not always perfectly selective. For example altering featural information by replacing the eyes and mouth with the ones from another face could also change their spatial relations (configural information) as mentioned in [10]. Rakover has pointed out that altering configuration by increasing the inter-eye distance could also induce a part-change, because the bridge of the nose might appears wider [11]. Such problems were avoided in our study by using scrambling and blurring procedures that allowed investigating the role of featural and configural information separately. The current study extends previous research using these manipulations (e.g. [12,13,14]) by ensuring that each procedure does effectively eliminate configural or featural processing.

Experiment 1: Unfamiliar Face Recognition

The first experiment investigated whether human observers store featural information *independent* of configural information. In the first condition configural information was eliminated by cutting the faces into their constituent parts and scrambling them. If the local information in parts (featural information) is encoded and stored, it should be possible to recognize faces above chance even if they are scrambled. In condition 2 the role of configural information was investigated. Previously learnt faces had to be recognized when they were shown as grayscale low-pass filtered versions. These image manipulations destroyed featural information while leaving the configural information intact. In a control condition we confirmed that performance is reduced to chance when faces are low-pass filtered and scrambled, thus showing that our image manipulations eliminate featural and configural information respectively and effectively.

Participants, Materials and Procedure

Thirty-six participants, ranging in age from 20 to 35 years voluntarily took part in this experiment. All were undergraduate students of psychology at Zurich University and all reported normal or corrected-to-normal vision.

The stimuli were presented on a 17” screen. The viewing distance of 1 m was maintained by a head rest so that the faces covered approximately 6° of the visual angle. Stimuli were created from color photographs of 10 male and 10 female undergraduate students from the University of Zurich who had agreed to be photographed and to have their pictures used in psychology experiments. All faces were processed with Adobe Photoshop, proportionally scaled to the same face width of 300 pixels and placed on a black background. These intact faces were used in the learning phase (Figure 1a).

The scrambled faces were created by cutting the intact faces into 10 parts, using the polygonal lasso tool with a 2 pixel feather. The number of parts was defined by a preliminary free listing experiment, in which 41 participants listed all parts of a face. The following parts were named by more than 80% of the participants and were used in this study: eyes, eyebrows, nose, forehead, cheeks, mouth, chin. Four different scrambling versions, which appeared randomly, were used. Each version was arranged so that no part was situated either in its natural position or in its natural relation to its neighboring part. The parts were distributed as close to each other as possible, in order to keep the image area approximately the same size as the whole faces (Figure 1b).

The control stimuli were created in three steps. First, all color information was discarded in the intact faces. Second, the faces were blurred using a Gaussian filter with a sigma of 0.035 of image width in frequency space, which was determined in pilot studies. The formula used to construct the filter in frequency space was $\exp(\frac{-f^2}{2\sigma^2})$. In the third step these blurred faces were cut and scrambled as described above. Figure 1c shows an example of the control faces.

The blurred stimuli were created by applying the low-pass filter determined in the control condition to greyscale versions of the intact faces (Figure 1d).

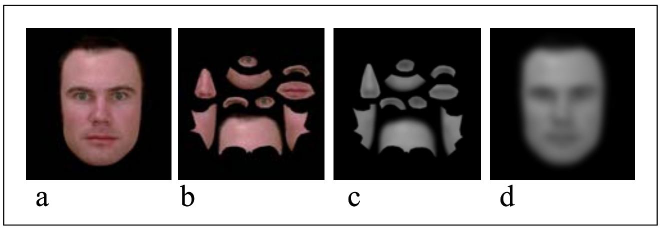

Fig. 1. Sample Stimuli. a) intact face, b) scrambled, c) scrambled-blurred, d) blurred face.

Participants were randomly assigned to one of three groups. Each group was tested in one experimental condition, either scrambled, scrambled-blurred, or blurred. Ten randomly selected faces served as target faces and the other 10 faces were used as distractors. In the learning phase the target faces were presented for ten seconds each. After each presented face the screen went blank for 1000 ms. Then the same faces were again presented 10 seconds each in the same order. The faces were presented in a pseudo-random order so that across participants no face appeared at the same position more than twice.

In the experimental phase, 20 faces were presented (10 targets and 10 distractors). Six random orders were created using the following constraints: within each random order no more than three target or distractor faces occurred on consecutive trials and between random orders no face appeared more than once on each position. The same random orders were used for all conditions.

Each trial started with a 1000 ms blank followed by a face. The participants were required to respond as fast and as accurately as possible whether the presented face was new (distractor) or whether it had been presented in the learning phase (target) by pressing one of two buttons on a response box. The assignment of buttons to responses was counterbalanced across participants.

Results and Discussion

Recognition performance was calculated using signal detection theory [15]. Face recognition performance was measured by calculating d' using an old-new recognition task [16]. This measure is calculated by the formula d' = z(H) – z(FA), whereas H denotes the proportion of hits and FA the proportion of false alarms. A hit was scored when the target button was pressed for a previously learned face (target) and a false alarm was scored when the target button was pressed for a new face (distractor). In the formula z denotes the z-transformation, i.e. H and FA are converted into z-scores (standard-deviation units). d' was calculated for each participant and averaged across each group (Figure 2, black bars).

One sample t-tests (one-tailed) were carried out in order to test the group means M against chance performance (i.e. d’ = 0). Faces were recognized above chance, even when they were cut into their parts, $M = 1.19$, $SD = 0.58$, $t(11) = 7.07$, $p < .001$. This result suggests that local part-based information has been encoded in the learning phase, which provided a useful representation for recognizing the scrambled versions in the testing phase. These findings are contradictory to the view that faces are only processed holistically [4,5,6,9]. The recognition of blurred faces was also above chance, $M = 1.67$, $SD = 0.82$, $t(11) = 7.044$, $p < .001$. The blur filter used did indeed eliminate all featural information since recognition was at chance when faces were blurred and scrambled, $M = -0.22$, $SD = 1.01$, $t(11) = -.75$, $p = .235$.

Taken together, these results provide clear evidence for the view that featural and configural information are both important sources of information in face recognition. Furthermore, the two processes do not appear to be arranged hierarchically, as the results show that featural and configural information can be encoded and stored independently of one another[1].

Experiment 2: Comparison of unfamiliar and familiar face recognition

The results of Experiment 1 challenge the hypothesis that faces are only processed holistically. At the same time our results suggest that for unfamiliar face recognition in humans separate representations exist for featural information and configural

[1] It is worth noting, however, that just because featural and configural processing *can* be recognized independently of one another, does not prove that the two don’t interact when both are available (e.g. [5])

information. The aim of Experiment 2 was to investigate whether the same is true for *familiar* face recognition. Moreover, by comparing recognition performance from Experiment 1 and Experiment 2 we addressed the question whether there is a shift in processing strategy from unfamiliar to familiar face recognition. Neuropsychological evidence suggests a dissociation between familiar face recognition and unfamiliar face matching [17,18], and experimental evidence suggests that familiar face recognition relies more heavily on the processing of inner areas of the face than does unfamiliar face recognition [19]. However, previous studies have found no evidence for a change in the balance between featural and configural processing as faces become more familiar [20,12]. Our study aimed to clarify this issue using a design that carefully controls the available featural and configural cues in the input image. Furthermore, in contrast to previous studies our study used the same faces in both experiments to eliminate other potential confounds with familiarity.

Participants, Materials and Procedure

Thirty-six participants ranging in age from 20 to 35 years took part in this experiment for course credits. All were undergraduate students of psychology at Zurich University and were familiar with the target faces. All reported normal or corrected-to-normal vision. The materials and procedure were the same as in Experiment 1. The stimuli were also the same, but all the targets were faces of fellow students and thus familiar to the participants. All distractor faces were unfamiliar to the participants.

Results and Discussion

The same analyses were carried out as in Experiment 1. Again, one-sample t-tests (one-tailed) revealed a significant difference from chance (i.e. d' > 0) for recognizing scrambled faces, $M = 2.19$, $t(11) = 4.55$, $p < .001$, and blurred faces, $M = 2.92$, $t(11) = 9.81$, $p < .001$. As in Experiment 1, scrambling blurred grayscale versions provided a control condition for testing whether the blur filter used did indeed eliminate all local part-based information. This was the case – faces could no longer be recognized when they were blurred and scrambled, $M = 0.19$, $t(11) = 0.94$, $p = .184$.

In short, the results of Experiment 2 replicated the clear effects from Experiment 1 and suggest an important role of local part-based and configural information in both unfamiliar and familiar face recognition. By comparing recognition performance from both experiments (Figure 2) we addressed the question to what extent familiar and unfamiliar face recognition differ quantitatively (e.g. generally a better performance when faces are familiar) or qualitatively (e.g. better performance for familiar faces using more accurate configural processing). To this end, a two-way analysis of variance (ANOVA) was carried out with the data from the scrambled and blurred conditions of Experiments 1 and 2 with familiarity (familiar vs. unfamiliar) and condition (scrambled vs. blurred) as between-subjects factors. There was a main effect of familiarity, $F(1,42) = 12.80$, $MSE = 13.48$, $p < .01$, suggesting that familiar faces are more reliably recognized than unfamiliar faces (quantitative difference). There was also a main effect of condition, $F\ (1,42) = 6.7$, $MSE = 7.05$, $p < .05$, indicating that blurred faces were better recognized than scrambled faces. The relative impact of blurring and scrambling did not differ between the two experiments, since there was no interaction between condition and familiarity, $F(1,42) = 1.02$, $MSE = 1.08$, $p = 0.32$. This results suggests that there are no qualitative differences between familiar and unfamiliar face recognition on the basis of configural and featural information. In both cases both types of information are of similar importance.

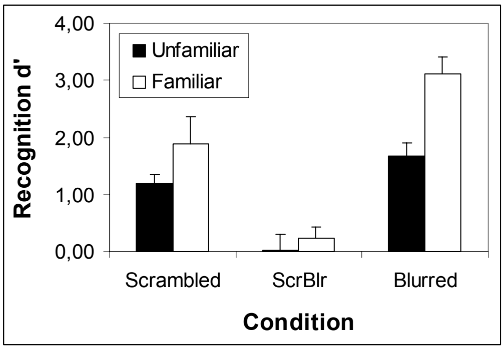

Fig. 2. Recognition performance in unfamiliar and familiar face recognition across the three different conditions at test. ScrBlr: scrambled and blurred faces. Error bars indicate standard errors of the mean.

General Discussion

In the present paper we investigated the role of local part-based information and their spatial interrelationship (configural information) using psychophysics. We found that human observers process familiar and unfamiliar faces by encoding and storing configural information as well as the local information contained in facial parts. These results challenge the assumption that faces are processed only holistically and suggest a greater biological plausibility for recent machine vision approaches in which local features and parts play a pivotal role (e.g. [1,2,3]).

Neurophysiological evidence supports part-based as well as configural and holisitic processing assumptions. In general, it has been found that cells responsive to facial identity are found in inferior temporal cortex while selectivity to facial expressions, viewing angle and gaze direction can be found in the superior temporal sulcus [21, 22]. For some neurons, selectivity for particular features of the head and face, e.g. the eyes and mouth, has been revealed [22,23,24]. Other groups of cells need the simultaneous presentation of multiple parts of a face and are therefore consistent with a more holistic type of processing [25,26]. Finally, Yamane et al. [27] have discovered neurons that detect combinations of distances between facial parts, such as the eyes, mouth, eyebrows, and hair, which suggest sensitivity for the spatial relations between facial parts (configural information).

In order to integrate the above mentioned findings from psychophysics, neurophysiology and computer vision we propose the framework depicted in Figure

3. Faces are first represented by a metric representation in primary visual areas corresponding to the perception of the pictorial aspects of a face. Further processing entails extracting local part-based information and spatial relations between them in order to activate featural and configural representations in higher visual areas of the ventral stream, i.e. face selective areas in temporal cortex[2]. In a recent study, repetition priming was used in order to investigate whether the outputs of featural and configural representations converge to the same face identification units [28]. Since priming was found from scrambled to blurred faces and vice versa we propose that the outputs of featural and configural representations converge to the same face identification units.

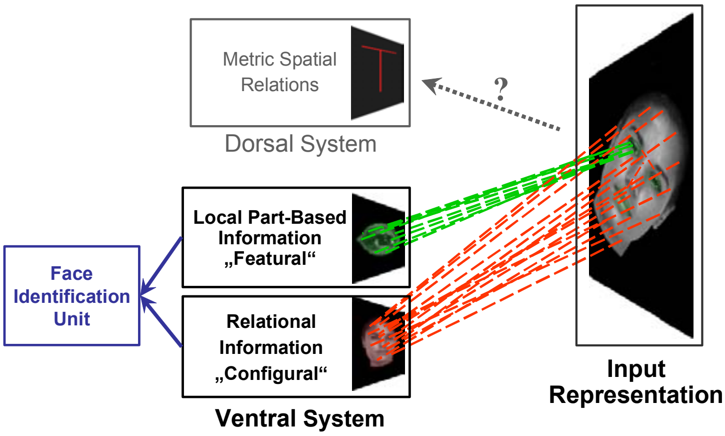

Fig. 3. Integrative model for unfamiliar and familiar face recognition.

References

1. Heisele, B., Serre, T., Pontil, M., Vetter, T., and Poggio, T. (2001). Categorization by learning and combining object parts. NIPS proceedings.
2. Lee, D.D., & Seung, S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401, 788-791.
3. Ullman, S., & Sali, E. (2000). Object Classification Using a Fragment-Based Representation. BMCV 2000. Lecture Notes in Computer Science, 1811, pp. 73-87. Berlin: Springer.
4. Tanaka J. W. & Farah, M. J. (1991). Second-order relational properties and the inversion-effect: Testing a theory of face perception. Perception & Psychophysics, 50, 367-372.
5. Tanaka, J. W. & Farah, M. J. (1993). Parts and wholes in face recognition. Quarterly Journal of Experimental Psychology, 79, 471-491.
6. Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? Journal of Experimental Psychology: Human Perception and Performance, 21, (3), 628-634.

[2] Although a role of the dorsal system in encoding of metric spatial relations has been proposed for object recognition it remains to be investigated, whether it does play a role for the processing of configural information in faces.

7. Lades, M., Vorbrüggen, J.C., Buhmann, J., Lange, J., Malsburg, v.d., C., Würtz, R.P., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers, 42, 300-311.
8. Wiskott, L., Fellous, J.M., Krüger, N., & von der Malsburg, C. (1997). Face Recognition by Elastic Bunch Graph Matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 775-779.
9. Biedermann, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. Philosophical Transactions of the Royal Society of London, B, 352, 1203-1219.
10. Rhodes, G., Brake, S., & Atkinson, A.P. (1993). What's lost in inverted faces? Cognition, 47, 25-57.
11. Rakover, S. S. (2002). Featural vs. configurational information in faces: A conceptual and empirical analysis. British Journal of Psychology, 93, 1-30.
12. Collishaw, S.M., Hole G.J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. Perception, 29, 893-910.
13. Davidoff, J., & Donnelly, N. (1990). Object superiority: a comparison of complete and part probes, Acta Psychologica, 73 1990 225–243.
14. Sergent J. (1985). Influence of task and input factors on hemispheric involvement in face processing. Journal of Experimental Psychology: Human Perception and Performance, 11(6), 846-61.
15. Green, D.M., & Swets, J.A. (1966). Signal detection theory and psychophysics. New York: Wiley.
16. McMillan, N.A., & Creelman, C.D. (1992). Detection theory: A user's guide. New York: Cambridge University Press.
17. Benton AL, 1980. The neuropsychology of facial recognition. American Psychologist, 35, 176-186.
18. Malone D.R., Morris H.H., Kay M.C., Levin, H.S., 1982. Prosopagnosia: a double dissociation between the recognition of familiar and unfamiliar faces. Journal of Neurology, Neurosurgery, and Psychiatry, 45, 820-822.
19. Ellis, H.D., Shepherd, J.W., Davies G.M., 1979. Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. Perception, 8, 431-439
20. Yarmey, A.D., 1971. Recognition memory for familiar "public" faces: Effects of orientation and delay. Psychonomic Science, 24, 286-288.
21. Hasselmo, M.E., Rolls, E.T., & Baylis, C.G. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. Experimental Brain Research, 32, 203-218.
22. Perret, D.I., Hietanen, J.K., Oram, M.W., & Benson, P.J. (1992). Organization and functions of cells in the macaque temporal cortex. Philosophical Transactions of the Royal Society of London, B, 335, 23-50.
23. Perret, D.I., Rolls, E.T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. Experimental Brain Research, 47, 329-342.
24. Perret, D.I., Mistlin, A.J., & Chitty, A.J. (1987). Visual neurones responsive to faces. Trends in Neuroscience, 10, 358-364.
25. Perret, D.I., & Oram, M.W. (1993). Image Vis. Comput., 11, 317-333.
26. Wachsmuth, E., Oram, M.W., & Perret, D.I. (1994). Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. Cerebral Cortex, 4, 509-522.
27. Yamane, S., Kaji, S., & Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? Experimental Brain Research, 73, 209-214.
28. Schwaninger, Lobmaier, & Collishaw (2002). Role and interaction of featural and configural processing in face recognition. Vision Sciences Society, 2nd annual meeting, Sarasota, Florida, May 10-15, 2002.