Anders, Ende, Junghofer, Kissler & Wildgruber (Eds.)
Progress in Brain Research, Vol. 156
ISSN 0079-6123

CHAPTER 18

Processing of facial identity and expression: a psychophysical, physiological and computational perspective

Adrian Schwaninger[1,2]*, Christian Wallraven[1], Douglas W. Cunningham[1] and Sarah D. Chiller-Glaus[2]

[1] *Department of Bülthoff, Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany*
[2] *Department of Psychology, University of Zurich, Zurich, Switzerland*

Abstract: A deeper understanding on how the brain processes visual information can be obtained by comparing results from complementary fields such as psychophysics, physiology and computer science. In this article, empirical findings are reviewed with regard to the proposed mechanisms and representations for processing identity and emotion in faces. Results from psychophysics clearly show that faces are processed by analyzing component information (eyes, nose, mouth, etc.) and their spatial relationship (configural information). Results from neuroscience indicate separate neural systems for recognition of identity and facial expression. Computer science offers a deeper understanding of the required algorithms and representations, and provides computational modeling of psychological and physiological accounts. An interdisciplinary approach taking these different perspectives into account provides a promising basis for better understanding and modeling how the human brain processes visual information for recognition of identity and emotion in faces.

Keywords: Face recognition, facial expression, interdisciplinary approach, psychophysics of face processing, computational modeling of face processing, face processing modules, component and configural processing

Introduction

Everyday object recognition is usually a matter of discriminating between quite heterogeneous object *classes* that differ with regard to their global shape, parts and other distinctive features such as color or texture. Face recognition, in contrast, relies on the discrimination of *exemplars* of a very homogenous category. According to Bahrick et al. (1975) we are able to recognize familiar faces with an accuracy of 90 % or more, even when some of these faces have not been seen for fifty years. Moreover, people identify facial expression very fast and even without awareness (see Leiberg & Anders, this volume). These abilities seem to be remarkably disrupted if faces are turned upside-down. Consider the pictures in Figure 1. Although this woman is a well-known celebrity, it is difficult to recognize her from the inverted photographs. One might detect certain differences between the two pictures despite the fact that both seem to have the same facial expression. Interestingly, after rotating this page 180 deg so that the two faces are upright, one can now easily identify the person depicted in these pictures and grotesque differences in the facial expression are revealed. This illusion has been discovered

*Corresponding author. Tel: +41-76-393-24-46; Fax: +49-7071-601-616;E-mail: adrian.schwaninger@tuebingen.mpg.de

DOI: 10.1016/S0079-6123(06)56018-2

Fig. 1. Thatcher illusion. When the photographs are viewed upside down (as above) it is more difficult to identify the person belonging to the pictures and the facial expressions seem similar. When the pictures are viewed right side up, it is very easy to identify the person depicted in these pictures and the face on the right appears highly grotesque.

by Thompson (1980). He used Margareth Thatcher's face which is why the illusion is known as the "Thatcher illusion". It was well known already by painters and Gestalt psychologists that face processing is highly dependent on orientation (e.g. Köhler, 1940). However, the finding that upside-down faces are *disproportionately* more difficult to recognize than other inverted objects, has been referred to as the *face inversion effect* and was first reported by Yin (1969).

Another interesting effect was discovered by Young et al. (1987). Composite faces were created by combining the top and bottom half of different faces. Figure 2a shows an example. If the two halves were aligned and presented upright, a new face resembling each of the two originals seemed to emerge. This made it difficult to identify the persons from either half. If faces were inverted or if the top and bottom halves were misaligned horizontally, then the two halves did not spontaneously fuse to create a new face, and the constituent halves remained identifiable.

Calder et al. (2000) used the same technique to investigate the processing of facial expressions. They prepared emotional face composites by aligning the top half of one expression (e.g., anger) with the bottom half of another (e.g., happiness) from the same person. When the face composites were aligned, a new facial expression emerged and participants were slower to identify the expression in either half of these composite images. However, this effect diminished when faces were misaligned or inverted, which parallels the composite effect for facial identity by Young et al. (1987). Interestingly, in an additional experiment Calder et al. found evidence for the view that the composite effects for identity and expression operate independently of one another.

These examples illustrate that information of parts and spatial relations are somehow combined in upright faces. In contrast, when faces are turned upside-down it seems that only the local part-based information is processed.

In this chapter, we discuss the representations and processes used in recognition of identity and facial emotion. We follow a cognitive neuroscience approach, by discussing the topic from a psychophysical, physiological and computational perspective. Psychophysics describe the relationship between stimuli in our external world and our internal representations. We first review the psychophysics literature on recognition of faces and facial expressions. Because our goal is to gain a deeper understanding of how our brain produces behavior, we discuss possible neural substrates of the representations and processes identified in neuroscience. Computer science, the third perspective, provides computational algorithms to solve certain recognition problems and the possibility of biologically plausible computer models.

Fig. 2. Aligned and misaligned halves of different identities (Margaret Thatcher and Marilyn Monroe). When upright (as above), a new identity seems to emerge from the aligned composites (left), which makes it more difficult to extract the original identities. This does not occur for the misaligned composite face (right). When viewed upside-down, the two identities do not fuse to a new identity.

Psychophysical perspective

Recognition of identity

Two main hypotheses have been proposed to explain the recognition of identity in faces: the holistic hypothesis and the component-configural hypothesis. According to the holistic hypothesis, upright faces are stored as unparsed perceptual wholes in which individual parts are not explicitly represented (Tanaka and Farah, 1991, 1993; Farah et al., 1995b). The main empirical evidence in favor of this view is based on a paradigm by Tanaka and Farah (1993). These authors argued that if face recognition relies on parsed representations, then single parts of a face, such as nose, mouth, or eyes, should be easily recognized even if presented in isolation. However, if faces are represented as unparsed perceptual wholes (i.e., holistically), then recognition of the same isolated parts should be more difficult. In their experiments, participants were shown a previously learned face together with a slightly different version in which one single part (e.g., nose or mouth) had been replaced. The task was to judge which face had been shown in the learning phase. The experiment was conducted in both a whole face condition and an isolated parts condition without facial context. In the isolated condition, face parts proved to be more difficult to recognize than in the whole face condition. However, when participants were trained to recognize inverted faces, scrambled faces, and houses no such advantage of context was found. Tanaka and Farah concluded that face recognition relies mainly on holistic representations, in contrast to the recognition of objects. While the encoding and matching of parts are assumed to be relatively orientation invariant (see also Biederman, 1987), holistic processing is thought to be very sensitive to orientation (see also Farah et al., 1995b; Biederman and Kalocsai, 1997).

The component-configural hypothesis is based on a qualitative distinction between component and configural information. The term component (or part-based, piecemal, feature-based, featural) information refers to those elements of a face that are perceived as parts of the whole (e.g., the eyes, mouth, nose, ears, chin, etc.). According to Bruce (1988), the term configural information (or configurational, spatial-relational, second-order relational information) refers to the "spatial interrelationship of facial features" (p. 38). Examples are the eye–mouth or intereye distance. Interestingly, these distances are overestimated by 30–40% (eye–mouth distance) and about 15% (intereye distance) in face perception (Schwaninger et al., 2003b). In practice, configural changes have been induced by altering the distance between

components or by rotating components (e.g., turning the eyes and mouth upside-down within the facial context like in the Thatcher illusion described above). According to the component-configural hypothesis, the processing of configural information is strongly impaired by inversion or plane rotation, whereas processing of component information is much less affected. There are now a large number of studies providing converging evidence in favor of this view (e.g., Sergent, 1984; Rhodes et al., 1993; Searcy and Bartlett, 1996; Leder and Bruce, 1998; Schwaninger and Mast, 2005; see Schwaninger et al., 2003a, for a review). These studies changed component information by replacing components (e.g., eyes of one person were replaced with the eyes of another person). Configural changes were induced by altering the distance between components (e.g., larger or smaller intereye distance). However, one possible caveat is that these types of manipulations often change the holistic aspects of the face and they are difficult to carry out selectively. For example, replacing the nose (component change) might change the distance between the contours of the nose and the mouth, which induces a configural change (Leder and Bruce, 1998, 2000). Moving the eyes apart (configural change) can lead to an increase in size of the bridge of the nose, which is a component change (see Leder et al., 2001). Such problems can be avoided by using scrambling and blurring procedures to reduce configural and component information independently (e.g., Sergent, 1985; Davidoff and Donnelly, 1990; Collishaw and Hole, 2000; Boutet et al., 2003). Schwaninger et al. (2002) extended previous research by ensuring that scrambling and blurring effectively eliminate configural and component information separately. Furthermore, in contrast to previous studies, Schwaninger et al. (2002) used the same faces in separate experiments on unfamiliar and familiar face recognition to avoid potential confounds with familiarity. In an old–new recognition paradigm it was found that previously learned intact faces could be recognized even when they were scrambled into constituent parts. This result challenges the assumption of purely holistic processing according to Farah et al. (1995b) and suggests that components are encoded and stored explicitly. In a second condition, the blur level was determined that made the scrambled versions impossible to recognize. This blur level was then applied to whole faces in order to create configural versions that by definition did not contain local featural information. These configural versions of previously learned intact faces could be recognized reliably. These results suggest that separate representations exist for component and configural information. Familiar face recognition was investigated in a second experiment by running the same conditions with participants who knew the target faces (all distractor faces were unfamiliar to the participants). Component and configural recognition was better when the faces were familiar, but there was no qualitative shift in processing strategy as indicated by the fact that there was no interaction between familiarity and condition (see Fig. 3).

Schwaninger et al. (2002, 2003a) proposed a model that allows integrating the holistic and component-configural hypotheses. Pictorial aspects of a face are contained in the pictorial metric input representation that corresponds to activation of primary visual areas. On the basis of years of experience, neural networks are trained to extract specific information in order to activate component and configural representations in the ventral visual stream. The output of these representations converges towards the same identification units. These units are holistic in the sense that they integrate component and configural information. Note that this concept of holistic differs from the original definition of Tanaka and Farah (1993) and Farah et al. (1995b), which implies that faces are stored as perceptual wholes without explicit representations of parts (component information).

The model by Schwaninger et al. (2002, 2003a) assumes that it is very difficult to mentally rotate a face as a perceptual whole (Rock, 1973, 1974, 1988; Schwaninger and Mast, 2005). When faces are substantially rotated from upright, they have to be processed by matching parts, which explains why information about their spatial relationship (configural information) is hard to recover when faces are inverted (for a similar view see Valentine and Bruce, 1988). Since face recognition depends on detecting subtle differences in configural

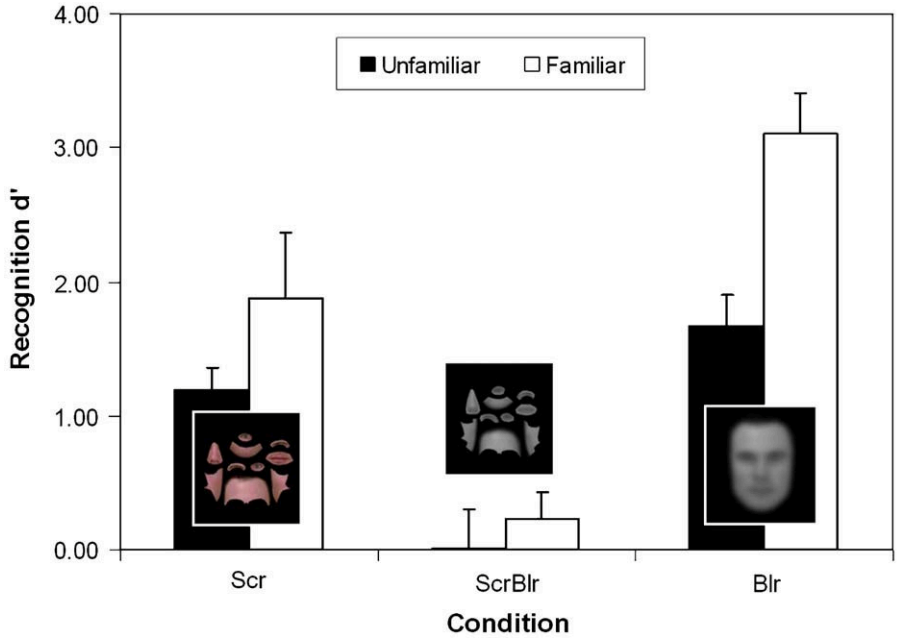

Fig. 3. Recognition performance in unfamiliar and familiar face recognition across three different conditions at test. Scr: scrambled; ScrBlr: scrambled and blurred; Blr: blurred. (Adapted with permission from Schwaninger et al., 2002.)

information, a large inversion effect is observed (Yin, 1969). Consistent with this view, Williams et al. (2004) suggested that inverted faces are initially processed by parts-based assessment before second-order relational processing is initiated. Sekuler et al. (2004) used response classification and found that the difference between the processing of upright and inverted faces was of quantitative, rather than of qualitative, nature, i.e., information was extracted more efficiently from upright faces than from inverted faces. This is also consistent with Schwaninger et al.'s model if one assumes that configural processing is not abruptly but gradually impaired by rotation (Murray et al., 2000; Schwaninger and Mast, 2005) and integrated with the output of component processing. Inverting the eyes and mouth within an upright face results in a strange activation pattern of component and configural representations that appears grotesque when upright, but not when upside-down (Thatcher illusion). This can be explained by the integrative model as follows: in an inverted Thatcher face, the components themselves are in the correct orientation which results in a relatively normal activation of component representations. The unnatural spatial relationship (changed configural information) is hard to perceive due to capacity limitations of an orientation normalization mechanism. As a consequence, the strange activation pattern of configural representations is reduced and the grotesque perception disappears. The composite face illusion can be explained on the basis of similar reasoning. Aligned upright face composites contain new configural information resulting in a new perceived identity. Inverting the aligned composites reduces the availability of configural information and it is easier to access the two different face identification units on the basis of component information alone.

Note that the model does not apply to processing component and configural information for gaze perception. As shown by Schwaninger et al. (2005), there is also an inversion effect on gaze perception. However, this effect is not due to impaired configural processing but due to orientation-sensitive processing of local component information in the eyes. This difference between inversion effects for recognition of identity versus

perceived eye gaze direction are consistent with separate functional systems for these tasks, which is consistent with physiological evidence discussed below.

In short, the model by Schwaninger et al. (2002, 2003a) allows the integration of the component configural hypothesis and holistic aspects of face processing relevant to recognition of identity. It can also explain striking perceptual effects such as the Thatcher illusion and the composite face illusion. Most importantly, it provides an integrative basis for understanding special characteristics of face recognition such as the specialization in upright faces and the sensitivity to configural information.

Recognition of expressions

The structure and perception of facial expressions have been subject to scientific examination since at least Duchenne's (1990) and Darwin's (1872) seminal work. The majority of these studies consisted of showing static photographs of expressions to observers and examining the relationship between statically visible deformations of the facial surface and the judgments made by the observers. It is, of course, clear that different facial areas are important for the recognition of different emotions (Hanawalt, 1944; Plutchik, 1962; Nummenmma, 1964; Bassili, 1979; Cunningham et al., 2005). For example, as mentioned above, Bassili (1979) used point-light faces to show that the upper portions of the face are important for some expressions, while the lower portions of the face are important for other expressions. Facial features also play differentiated roles in other aspects of facial expression processing, such as the perception of sincerity. For example, according to Ekman and Friesen (1982), a true smile of enjoyment, which Ekman refers to as a Duchenne smile, has a characteristic mouth shape as well as specific wrinkles around the eyes. Faked expressions of enjoyment, in contrast, contain just the mouth information. Furthermore, Ekman and Friesen (1982) have shown that deceptive expressions of enjoyment appear to have different temporal characteristics than spontaneous ones.

Given the general preoccupation with the role of featural information in the recognition of facial expressions, it should not be surprising that the vast majority of descriptive systems and models of facial expressions are explicitly part-based (Frois-Whittmann, 1930; Frijda and Philipszoon, 1963; Leventhal and Sharp, 1965; Ekman and Friesen, 1978; Izard, 1979; Tronick et al., 1980; Essa and Pentland, 1994; Ellison and Massaro, 1997). Perhaps the most widely used methods for parametrizing the high-dimensional space of facial expressions is the facial action coding system (or FACS; Ekman and Friesen, 1978), which segments the visible effects of facial muscle activity and rigid head motion into "action units." Combinations of these action units can then be used to describe different expressions. It is important to note that FACS was designed as a system for describing the elements of photographs of facial expressions. It is not a model of facial expression processing and makes no claims about which elements go together to produce different expressions (Sayette et al., 2001).

Massaro and colleagues proposed a parts-based model of perception (the fuzzy logical model of perception or FLMP) in which the features are independently processed and subsequently integrated. The model makes specific claims about how the featural information is processed and integrated, and thus makes clear predictions about the perception and categorization of facial expressions. In one study, Ellison and Massaro (1997) used computer graphics animation techniques to produce static facial expressions where either (a) the mouth shape was parametrically varied, (b) the eyebrow shape was parametrically varied, or (c) both were independently parametrically varied. The faces were shown to a number of observers, who were asked if the expression in the photographs was that of happiness or anger. Ellison and Massaro found that both features (eyebrow position and mouth position) affected the participants' judgments, and that the influence of one feature was more prominent when the other feature was neutral or ambiguous. Moreover, the FLMP captured patterns in the data better than either holistic models or a straight-forward additive model based on recognition rates of the individual features.

Elison and Massaro consequently claimed that the perceptual system must be using featural information in the recognition process and cannot be employing a purely holistic approach. These results are consistent with the finding that the aligned combination of two different emotions leads to decreased recognition performance (Calder et al., 2000).

Just as is true for the processing of facial identity, the separate roles of component-configural and holistic information have been discussed within the context of facial expressions. There are at least two models that integrate holistic information (Izard et al., 1983; White, 2000). White (2000) proposed a "hybrid model," according to which expression recognition is part-based on one hand and holistic in the sense of undecomposed wholes on the other hand.

Several researchers have examined the role of temporal information in the perception and recognition of expressions (Bassili, 1978, 1979; Bruce, 1988; Edwards, 1998; Kamachi et al., 2001). Kamachi et al., for example, manipulated the velocity in which a neutral face turned into an emotional one. They found that happiness and surprise were better recognized from fast sequences, sadness was better recognized from slow sequences, and angriness was best recognized at medium speed. This indicates that different expressions seem to have a characteristic speed or rate of change. In an innovative study, Edwards (1998) presented participants with photographs of individual frames from a video sequence of a dynamic expression, in a scrambled order, and asked participants to place the photographs in the correct order. Participants were remarkably accurate in their reconstructions, showing a particularly strong sensitivity to the temporal characteristics in the early phases of an expression. Interestingly, participants performed better when asked to complete the task with extremely tight time constraints than when given unlimited time, from which Edwards concluded that conscious strategies are detrimental to this task. He further concluded that the results show that humans do encode and represent temporal information about expressions.

In sum, it is clear that different expressions require different features to be recognized, that one can describe expressions in terms of their features and configuration, and that dynamic information is represented in the human visual system and is important for various aspects of facial expression processing. Moreover, simple, purely holistic models do not seem to describe the perception of facial expressions very well.

Dynamic information

The vast majority of research on the perception of faces has tended to focus on the relatively stable aspects of faces. This has consequently led to a strong emphasis on static facial information (i.e., information that is available at any given instant, such as eye color, distance between the eyes, etc.). In general, however, human faces are not static entities. Humans are capable of moving their faces in a wide variety of ways, and they do so for an astonishingly large number of reasons. Recent advances in technology, however, have allowed researchers to begin examining the role of motion in face processing.

Before one can determine what *types* of motion are used in the recognition of faces and facial expressions (i.e., the dynamic features), one must determine if motion plays any role at all. To this end, it has been clearly established that dynamic information can be used to recognize identity (Pike et al., 1997; Bruce et al., 1999, 2001; Lander et al., 1999, 2001; Knappmeyer et al., 2003) and to recognize expressions (Bassili, 1978, 1979; Humphreys et al., 1993; Edwards, 1998; Kamachi et al., 2001; Cunningham et al., 2005; Wallraven et al., 2005a). Overall, the positive influence of dynamic information is most evident when static information is degraded.

It is difficult, if not impossible, to present dynamic information without static information. Thus, Pike et al. (1997) and Lander and colleagues performed a series of control experiments to ensure that the apparent advantage moving faces have over static faces is due to information that is solely available over time (i.e., dynamic information). One might, for example, describe dynamic sequences as a series of static snapshots. Under such a description, the advantage of dynamic stimuli would not lie with dynamic information,

but with the fact that a video sequence has more static information (i.e., it has *supplemental information* provided by the different views of the face). To test this hypothesis, Lander et al. (1999) asked participants to identify a number of famous faces. The faces were presented in three different formats. In one condition, participants saw a nine-frame video sequence. In a second condition, participants saw all nine frames at once, arranged in an ordered array. In the final condition, the participants saw all nine frames at once in a jumbled array. Lander et al. found that the faces were better recognized in the video condition than in the either of the two static conditions, and that the performances in the two static conditions did not differ from one another. Thus, the reason why video sequences are recognized better is not simply that they have more snapshots.

To test whether the advantage is due to motion in general or due to some specific type of motion, Lander and Bruce (2000) and Pike et al. (1997) presented a video where the images were in a random order. Note that such sequences have information about the motion, but this motion is random (and does not occur in nature). It was found that identity is more accurately recognized in normal sequences that in random sequences, implying that it is not just the presence of motion that is important, but the specific, naturally occurring motion that provides the advantage. Further, it was found that reversing the direction of motion (by playing the sequence backwards) decreases recognition performance, suggesting that the temporal direction of the motion trajectories is important (Lander and Bruce, 2000). Finally, by changing the speed of a motion sequence (e.g., by playing parts or all of a video sequence too fast or slow), the researchers showed that the specific tempo and rhythm of motion is important for face recognition (Lander and Bruce, 2000).

In a perhaps more direct examination of the role of motion, Bassili (1978, 1979) used Johonnson point-light faces as stimuli (see Johonnson, 1973, for more information on point-light stimuli). More specifically, the face and neck of several actors and actresses were painted black and then covered with approximately 100 white spots. These actors and actresses were then recorded under low light conditions performing either specific expressions (happy, sad, surprise, disgust, interest, fear, and anger) or any facial motion the actor/actress desired. They kept their eyes closed during the recording sessions. Thus, in the final video sequence, all that was visible were the 100 white points. Each participant saw a single display, either as a single static snapshot or a full video recording of one expression, and was asked to describe what they saw. The collection of points was recognized as being a face more often in the dynamic conditions than in the static conditions (73% vs. 22% of the time, respectively). Additionally, the sequences were recognized as containing a face slightly more often for the free-form motion conditions than for the expression conditions (55% vs. 39%, on average, respectively). In a second experiment, the actors and actresses were again recorded while performing the various emotions. In the first set of recordings, they again wore the black makeup and white spots. The second set of recordings was made without makeup. Participants were asked to identify the expression using a forced choice task. Overall, faces were recognized more often in the full-face condition than in the dots-only condition (65% vs. 33% correct responses, respectively). Critically, the percentage of correct responses in the point-light condition (33%) is significantly higher than expected by chance, suggesting that the temporal information is sufficient to recognize expressions. Basilli (1979) went on to examine the role of upper versus lower internal facial motion for the recognition of expressions and found that different facial areas were important for different expressions.

It remains unclear as exactly what the appropriate dynamic features are. One traditional way of describing motion is to separate it into rigid and nonrigid motions (see, e.g., Gibson, 1957, 1966; Roack et al., 2003). Rigid face motion generally refers to the rotations and translations of the entire head (such as the one which occurs when someone nods his/her head). Nonrigid face motion, in contrast, generally refers to motion of the face itself, which consists mostly of nonlinear surface deformations (e.g., lip motion, eyebrow motion). Most naturally occurring face-related motion contains both rigid and nonrigid motion. Indeed, it is very difficult for humans to produce facial (i.e.,

nonrigid) motion without moving their head (rigid motion), and vice versa. Thus, it should not be surprising that few studies have systematically examined the separate contributions of rigid and nonrigid face motions. Pike et al. (1997) conducted one of the few studies to explicitly focus on the contribution of rigid motion. They presented a 10 s clip of an individual rotating in a chair through a full 360° (representing a simple change in relative viewpoint). They found higher identity recognition performance in dynamic conditions than in static conditions. Christie and Bruce (1998), in contrast, presented five frames of a person moving his/her head up and down (e.g., representing social communication — a nod of agreement) and found no difference between static and dynamic conditions. They suggest that the apparent conflict between the two studies comes from the type of rigid head motion: viewpoint change versus social signal. Munhall et al. (2004) focused explicitly on the role of rigid head motion in communication and showed in an elegant study that the specific pattern of rigid head motion that accompanies speech can be used to disambiguate the speech signal when the audio is degraded. Hill and Johnson (2001) used facial animations to show that rigid head motion is more useful than nonrigid motion for identity recognition and that nonrigid motion was more useful than rigid motion in recognizing the gender of an individual.

In sum, it is clear that some form of facial information is available only over time, and that it plays an important role in the recognition of identity, expression, speech, and gender. Moreover, at least several different types of motion seem to exist, they play different roles, and a simple rigid/nonrigid dichotomy is neither sufficient nor appropriate to describe these motions. Additional research is necessary to determine what the dynamic features for face processing are.

Physiological perspective

Face-selective areas — evidence from neuroscience

At least since the discovery of the face inversion effect (Yin, 1969) it has been discussed whether a specific area for the processing of faces exists in the human brain. Neuropsychological evidence for specialization has been derived from prosopagnosia, a deficit in face identification following inferior occipitotemporal lesions (e.g., Damasio et al., 1982; for a review see DeRenzi, 1997). There have been a few reports of prosopagnostic patients in which object recognition seemed to have remained intact (e.g., McNeil and Warrington, 1993; Farah et al., 1995a; Bentin et al., 1999). Prosopagnosia has been regarded as a face-specific deficit that does not necessarily reflect a general disorder in exemplar recognition (e.g., Henke et al., 1998). Consistent with this view, patients who suffered from associative object agnosia have been reported, while their face identification remained unaffected (e.g., Moscovitch et al., 1997). Such a double dissociation between face and object recognition would imply that the two abilities are functionally distinct and anatomically separable. However, on the basis of methodological concerns, some authors have doubted whether face recognition can really be dissociated from object recognition based on current literature on prosopagnosia (e.g., Gauthier et al., 1999a; see also Davidoff and Landis, 1990).

Evidence for the uniqueness of face processing has also been derived from event-related potential (ERP) and magnetoencephalographic (MEG) studies. A response component called the N170 (or M170 in MEG) occurring around 170 ms after stimulus onset is usually twice as large for face stimuli when compared to other control stimuli such as hands, houses, or animals (e.g., Bentin et al., 1996; Liu et al., 2002). However, the debate on whether such activation is unique for faces or whether it represents effects of expertise that are not specific to face processing is still ongoing (for recent discussions see, e.g., Rossion et al., 2002; Xu et al., 2005).

In functional brain imaging, several areas have been identified as being of special importance for the processing of faces (see Haxby et al., 2000, for a review). These involve a region in the lateral fusiform gyrus, the superior temporal sulcus (STS), and the "occipital face area" (OFA; Gauthier et al., 2000a). All areas have been

identified bilaterally, albeit with a somewhat stronger activation in the right hemisphere. The face-selective area in the fusiform gyrus has been referred to as the "fusiform face area" (FFA) by Kanwisher et al. (1997). While FFA activation has been related to facial identity, the STS in humans reacts particularly to changing aspects of faces with social value, such as expression, direction of gaze, and lip movement (e.g., Puce et al., 1998; Hoffman and Haxby, 2000). In a recent functional magnetic resonance imaging (fMRI) study using adaptation (reduction of brain activity due to repetitive stimulus presentation), Andrews and Ewbank (2004) investigated differences in face processing by the FFA versus the STS. Activity in the FFA was reduced over time by stimuli having the same identity. Adaptation was dependent on viewpoint but not on size changes. The STS showed no adaptation to identity but an increased response when the same face was shown with a different expression or from different viewpoints. These results suggest a relatively size-invariant neural representation in FFA for recognition of facial identity, and a separate face-selective region in STS involved in processing changeable aspects of a face such as facial expression. OFA and inferior occipital gyrus seem to be associated with early structural encoding processes; they are primarily sensitive to sensory attributes of faces (Rotshtein et al., 2005). Rossion et al. (2003) obtained results in an fMRI study suggesting that OFA and FFA might be functionally associated: PS, a patient suffering from severe prosopagnosia due to lesions in the left middle fusiform gyrus and the right inferial occipital cortex, performed poorly in a face-matching task despite normal activation of the intact right FFA. Rossion et al. thus concluded that the FFA alone does not represent a fully functional module for face perception, but that for normal face processing intact OFA and FFA in the right hemisphere with their re-entrant integration are necessary. Yovel and Kanwisher (2005) came to a different conclusion. They correlated the behavioral performance in a face-matching task of upright and inverted faces with the neuronal responses to upright and inverted faces in the three regions: FFA, STS, and OFA. It was found that only the FFA showed a difference in activity between upright and inverted faces. This can be interpreted as functional dissociation between FFA and the other cortical regions involved in face processing. The authors also concluded that the FFA appears to be the main neurological source for the behavioral face inversion effect originally reported by Yin (1969). The latter, however, is not exclusive to faces. In a behavioral study, Diamond and Carey (1986) found comparable inversion effects for faces and side views of dogs when dog experts were tested. Subsequent behavioral and imaging studies using recognition experiments with trained experts and artificial objects ("Greebles") as well as bird and car experts with bird and car images provided further evidence in favor of a process-specific interpretation rather than a domain-specific interpretation (Gauthier et al., 1999b, 2000a). According to their view ("expertise hypothesis"), FFA activity is related to the identification of different classes of visual stimuli if they share the same basic configuration and if substantial visual expertise has been gained. The question on whether FFA activity is domain or process specific is being debated since several years now. It is beyond the scope of this chapter to review this ongoing debate but for an update on the current status see, for example, Downing et al. (2005), Xu (2005), Bukach et al. (in press), Kanwisher and Yovel (in press). Nevertheless, it should be noted that activation in face-selective regions of the fusiform area is not exclusive to faces. Significant responses to other categories of objects have been found in normal subjects, for example, for chairs, houses, and tools (Chao et al., 1999; Ishai et al., 1999, 2000; Haxby et al., 2001). Moreover, it has also been shown that face-selective regions in the fusiform area can be modulated by attention, emotion, and visual imagery, in addition to modulation by expertise as mentioned above (e.g., O'Craven et al., 1999; Vuilleumier et al., 2001; Ishai et al., 2002). In recent years, substantial progress has been made regarding models on how different brain areas interact in processing information contained in faces. Three main accounts are summarized in the following section.

Cognitive neuroscience models of face processing

The model by Bruce and Young (1986) is one of the most influential accounts in the psychological face processing literature. This framework proposes parallel routes for recognizing facial identity, facial expression, and speech-related movements of the mouth. It is a rather functional account since Bruce and Young did not provide specifics regarding the neural implementation of their model. The recent physiological framework proposed by Haxby et al. (2000) is consistent with the general conception proposed by Bruce and Young. According to Haxby et al.'s model, the visual system is hierarchically structured into a core and an extended system. The core system comprises three bilateral regions in occipitotemporal visual extrastriate cortex: inferior occipital gyrus, lateral fusiform gyrus, and STS. Their function is the visual analysis of faces. Early perception of facial features and early structural encoding processes are mediated by processing in inferior occipital gyrus. The lateral fusiform gyrus processes invariant aspects of faces as the basis for the perception of unique identity. Changeable properties such as eye gaze, expression, and lip movement are processed by STS. The representations of changeable and invariant aspects of faces are proposed to be independent of one another, consistent with the Bruce and Young model. The extended system contains several regions involved in other cognitive functions such as spatially directed attention (intraparietal sulcus), prelexical speech perception (auditory cortex), emotion (amygdala, insula, limbic system), and personal identity, name, and biographical information (anterior temporal region).

The model of Haxby et al. has been taken as a framework for extension by O'Toole et al. (2002). By taking into account the importance of dynamic information in social communication, they further explain the processing of facial motion. In their system, dynamic information is processed by the dorsal stream of face recognition and static information is processed by the ventral stream. Two different types of information are contained in facial motion: social communication signals such as gaze, expression, and lip movements, which are forwarded to the STS via the middle temporal (MT) area; and person-specific motion ("dynamic facial signatures"). O'Toole et al. suggest that the latter type of information is also processed by the STS, representing an additional route for familiar face recognition. This model is in accordance with the *supplemental information hypothesis* that claims that facial motion constitutes additional information to static information. According to O'Toole et al., structure-from-motion may also support face recognition by communication between the ventral and the dorsal streams. For instance, the structural representation in FFA could be enhanced by input from MT. Thus, the model also integrates the *representation enhancement hypothesis*.

In a detailed review of psychological and neural mechanisms, Adolphs (2002) provides a description of the processing of emotional facial expressions as a function of time. The initial stage provides automatic fast perceptual processing of highly salient stimuli (e.g., facial expressions of anger and fear). This involves the superior colliculus and pulvinar, as well as activation of the amygdala. Cortical structures activated in this stage are V1, V2, and other early visual cortices that receive input from the lateral geniculate nucleus of the thalamus. Then, a more detailed structural representation of the face is constructed until about 170 ms. This processing stage involves the fusiform gyrus and the superior temporal gyrus, which is consistent with Haxby et al.'s core system. Dynamic information in the stimulus would engage MT, middle superior temporal area, and posterior parietal visual cortices. Recognition modules for detailed perception and emotional reaction involve Haxby et al.'s extended system. After 300 ms conceptual knowledge of the emotion signaled by the face is based on late processing in the fusiform and superior temporal gyri, orbitofrontal and somatosensory cortices, as well as activation of the insula.

The assumption of separate processes for facial identity and facial expression is supported by a number of studies. Neuropsychological evidence suggests a double dissociation; some patients show impairment in identity recognition but normal emotion recognition, and other patients show

intact identity recognition but impaired emotion recognition (for reviews see Damasio et al., 1982, 1990; Wacholtz, 1996; Adolphs, 2002). In a recent study, Winston et al. (2004) revealed dissociable neural representations of identity and expression using an fMRI adaptation paradigm. They found evidence for identity processing in fusiform cortex and posterior STS. Coding of emotional expression was related to a more anterior region of STS. Bobes et al. (2000) showed that emotion matching resulted in a different ERP scalp topography compared to identity matching. In another ERP study, Eimer and Holmes (2002) investigated possible differences in the processing of neutral versus fearful facial stimuli. They found that the N170, which is related to structural encoding of the face in processing identity, did occur in both the neutral and the fearful conditions. This indicates that structural encoding is not affected by the presence of emotional information and is also consistent with independent processing of facial expression and identity. However, results from other studies challenge the assumption of completely independent systems. DeGelder et al. (2003) found that subjects suffering from prosopagnosia performed much better when faces showed emotions than when they depicted a neutral expression. With normal subjects, the case was the opposite. DeGelder et al. assume that the areas associated with expression processing (amygdala, STS, parietal cortex) have a modulatory role in face identification. Their findings challenge the notion that different aspects of faces are processed independently (assumption of dissociation) and only after structural encoding (assumption of hierarchical processing). Calder and Young (2005) share a similar view. They argue that a successful proof of the dissociation of identity and expression would require two types of empirical evidence. First, patients with prosopagnosia but without any impairment in facial expression recognition. Second, intact processing of facial identity and impaired recognition of emotion without impairment of other emotional functions. On the basis of their review the authors conclude that such clear patterns have not been revealed yet. The reported selective disruption of facial expression recognition would rather reflect an impairment of more general systems than damage (or impaired access) to visual representations of facial expression. The authors do not completely reject the dissociation of identity and expression, but they suggest that the bifurcation takes place at a much later stage than that proposed by the model of Haxby et al., namely only after a common representational system. This alternative approach is supported by computational modeling studies using principal component analysis (PCA; see next section). A critical problem of these approaches, however, is that they rely on a purely holistic processing strategy of face stimuli, which in light of the previously discussed behavioral evidence seems not plausible.

As discussed in the previous section, there is a growing number of studies in the psychophysics literature that clearly suggest an important role of both component and configural information in face processing. This is supported by neurophysiological studies. 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 STS (Hasselmo et al., 1989; Perret et al., 1992). For some neurons, selectivity for particular features of the head and face, e.g. the eyes and mouth, has been revealed (Perret et al., 1982, 1987, 1992). Other groups of cells need the simultaneous presentation of multiple parts of a face, which is consistent with a more holistic type of processing (Perret and Oram, 1993; Wachsmuth et al., 1994). Yamane et al. (1988) 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).

Although they are derived from different physiological studies, the three models by Haxby, O'Toole et al., and Adolphs share many common features. Nevertheless, it seems that some links to behavioral and physiological studies are not taken up in these models. As discussed above, the concept of component and configural processing seems to be a prominent characteristic of face processing. The models, however, do not make this processing step explicit by specifying at which stage this information is extracted. Furthermore,

the distributed network of brain regions involved in the processing of face stimuli has so far not been characterized in terms of the *features* that are processed in each region — how does a face look like for the amygdala, for example? Some of these questions may be answered in connection with a closer look at the computational properties of face recognition. In the next section, we therefore present a brief overview of computational models of identity and expression recognition.

Computational perspective

Since the advent of the field of computer vision (Marr, 1982), face recognition has been and continues to be one of its best-researched topics with hundreds of papers being published each year in conferences and journals. One reason for this intense interest in face recognition is certainly due to the growing range of commercial applications for computational face recognition systems — especially in the areas of surveillance and biometrics, but increasingly also in other areas such as human–computer interaction or multimedia applications. Despite these tremendous efforts, however, even today there exists no single computational system that is able to match human performance — both in terms of recognition discriminability and in terms of generalization to new viewing conditions including changes in lighting, pose, viewing distance, etc. It is especially this fact that has led to a growing interest of the computer vision community in understanding and applying the perceptual, cognitive, and neurophysiological issues underlying human performance. Similar statements could be made about the area of automatic recognition of facial expressions — the critical difference being that commercial interest in such systems is less than in systems that can perform person identification. Nevertheless, the area of expression recognition continues to be a very active topic in computer vision because it deals with the temporal component of visual input: how the face moves and how computers might be able to map the space of expressions are of interest for computer vision researchers leading to potential applications in, for example, human–computer interaction in which the actions of the user have to be analyzed and recognized in a temporal context.

As the previous sections have shown, however, apart from having commercial prospects, techniques developed by the computer vision community also have wide uses in cognitive research: by analyzing and parametrizing the high-dimensional space of face appearances, for example, researchers gain access to a high-level, statistical description of the underlying visual data. This description can then be used to design experiments in a well-defined subspace of facial appearance (for a review of face spaces see Valentine, 1995). A well-known example consists of the PCA of faces that defines prototypes in a face space (Leopold et al., 2001). The same holds true in the case of facial expressions as the amount of spatio-temporal data quickly becomes prohibitive in order to conduct controlled experiments at a more abstract level that goes beyond mere pixels. A recent study that has used advanced computer vision techniques to manipulate components of facial expressions is the study by Cunningham et al. (2005). In the following, we will briefly review the main advances and approaches in both the area of identity recognition and recognition of expressions (see Li and Jain, 2004, for further discussion).

As a first observation, it is interesting to note that both identity and expression recognition in computer vision follow the same basic structure: in the first step, the image is scanned in order to find a face — this stage is usually called *face detection* and can also encompass other tasks such as estimating the pose of the face. As a result of space restrictions, we will not deal with face detection explicitly — rather, the reader is referred to Hjelmas and Low (2001). Interestingly, the topic of face detection has received relatively little attention in cognitive research so far (see, e.g., Lewis and Edmonds, 2003) and needs to be further explored. Following face detection, in a second step the image area that comprises the face is further analyzed to extract discriminative *features* ranging from holistic approaches using the pixels themselves to more abstract approaches extracting the facial components. Finally, the extracted features are compared to a database of stored identities or

expressions in order to recognize the person or their expression. This comparison can be done by a range of different classification schemes from simple, winner-take-all strategies to highly complex algorithms from machine learning.

Research in face recognition can be roughly divided into three areas following the type of information that is used to identify the person in the feature extraction step: (1) Holistic approaches use the full image pixel information of the area subtended by the face. (2) Feature-based approaches try to extract more abstract information from the face area ranging from high-contrast features to semantic facial features. (3) Hybrid systems combine these two approaches.

The earliest work in face recognition focused almost exclusively on high-level, feature-based approaches. Starting in the 1970s, several systems were proposed which relied on extracting facial features (eyes, mouth, and nose) and in a second step calculated two-dimensional geometric properties of these features (Kanade, 1973). Although it was shown that recognition using only geometric information (such as distances between the eyes, the mouth, etc.) was computationally effective and efficient, the robust, automatic extraction of such high-level facial features has proven to be very difficult under general viewing conditions (Brunelli and Poggio, 1993). One of the most successful face recognition systems based on local image information therefore used much simpler features that were supplemented by rich feature descriptors: in the elastic bunch-graph matching approach, a face image is represented as a collection of nodes which are placed in a regular grid over the face. Each of these nodes carries so-called ''jets'' of Gabor-filter responses, which are collected over various scales and rotations. This representation is very compact yet has proven to be very discriminative, therefore enabling good performance even under natural viewing conditions (Wiskott et al., 1997). It is interesting to note this system's similarity to the human visual system (see Biederman and Kalocsai, 1997) as the Gabor filters used closely resemble the receptive field structure found in the human cortex. The advantage of such low-level features as used also in later recognition systems lies in their conceptual simplicity and compactness.

In the early 1990s, Turk and Pentland (1991) developed a holistic recognition system called ''eigenfaces,'' which used the full pixel information to construct an *appearance-based* low-dimensional representation of faces. This approach proved to be very influential for computer vision in general and inspired many subsequent recognition algorithms. Its success is partially due to the fact that natural images contain many statistical redundancies. These can be exploited by algorithms such as PCA by building lower-dimensional representations that capture the underlying information contained in, for example, the space of identities given a database of faces. The result of applying PCA to such a database of faces is a number of eigenvectors (the ''eigenfaces'') that encode the main statistical variations in the data. The first eigenvector is simply the average face and corresponds to the prototype face used in psychology. Recognition of a new face image is done by projecting it into the space spanned by the eigenvectors and looking for the closest face in that space. This general idea of a face space is shared by other algorithms such as linear discriminant analysis (LDA; Belhumeur et al., 1997), independent component analysis (ICA; Bartlett et al., 2002), non-negative matrix factorization (NMF; Lee and Seung, 1999), or support vector machines (SVMs; Phillips, 1999). The main difference between these algorithms lies in the statistical description of the data as well as in the metrics used to compare different elements of the face space: PCA and LDA usually result in holistic descriptions of the data where every region of the face contributes to the final result; ICA and NMF can yield more sparse descriptors with spatially localized responses; SVMs describe the space of face identities through difficult-to-recognize face exemplars (the support vectors) rather than through prototypical faces as PCA does. In terms of metrics, there are several possibilities ranging from simple Euclidean distances to weighted distances, which can take the statistical properties of the face space into account.

The advantage of PCA (and other holistic approaches) in particular is that it develops a generative model of facial appearance which enables it, for example, to *reconstruct* the appearance of a noisy or occluded input face. An extreme example

of this is the morphable model by Blanz and Vetter (1999, 2003), which does not work on image pixels but works on *three-dimensional* (3D) data of laser scans of faces. Because of their holistic nature, however, all of these approaches require specially prepared training and testing data with very carefully aligned faces in order to work optimally.

Given the distinction between local and holistic approaches, it seems natural to combine the two into hybrid recognition architectures. Eigenfaces can of course be extended to "eigenfeatures" by training facial features instead of whole images. Indeed, such systems have been shown to work much better under severe changes in the appearance of the face such as due to occlusion by other objects or make-up (see Swets and Weng, 1996). Another system uses local information extracted from the face to fit a holistic shape model to the face. For recognition, not only holistic information is used, but also local information from the contour of the face (Cootes et al., 2001). Finally, in a system proposed by Heisele et al. (2003), several SVMs are trained to recognize facial features in an image, which are then combined into a configuration of features by a higher-level classification scheme. Again, such a scheme has been shown to outperform other, purely holistic, approaches.

Recently, several rigorous testing schemes have been proposed to evaluate the various methods proposed by computer vision researchers: the FERET (FacE REcognition Technology) evaluations in 1994–1996 and three face recognition vendor tests in 2000, 2002, and 2005 (see http://www.frvt.org). Although these evaluations have shown that performance has increased steadily over the past years, several key challenges still need to be addressed before face recognition systems can become as good as their human counterpart:

Robust extraction of facial features: Although face detection is possible with very high accuracy, this accuracy is usually achieved by analyzing large databases of face and nonface images to extract statistical descriptors. These statistical descriptors usually do *not* conform to meaningful face components — components that humans rely on to detect faces.

Tolerance to changes in viewing conditions: Existing databases often contain only frontal images or a single illumination. In the real world, however, changes in illumination and pose are rather frequent — usually several changes occur at the same time.

Dealing with large amounts of training data: Connected to the previous point, typical recognition systems need massive amounts of training data to learn facial appearance under various viewing conditions. While this is reminiscent of the extensive experience humans acquire with faces, this represents a challenge for the statistical algorithms as the relevant discriminative information has to be extracted from the images.

Taking into account the context: Faces seldom appear without context — in this case context can mean a particular scene, evidence from other modalities, or the fact that faces are part of the human body and therefore co-occur. This information could be used not only for more reliable face detection, but could also assist in recognition tasks. Interestingly, context effects are also not well studied in the behavioral literature.

Dealing with spatio-temporal data: Even though humans can recognize faces from still images, we generally experience the world dynamically (see Section "Dynamic information" above). Although first steps have been made (Li and Chellappa, 2001), the full exploitation of this fact remains to be done.

Characterizing performance with respect to humans: Although the claim that current systems do not yet reach human performance levels is certainly valid, there has been relatively little research on trying to relate computational and human performance in a *systematic* manner (examples include the studies by Biederman and Kaloscai, 1997; O'Toole et al., 2000; Wallraven et al., 2002, 2005b; Schwaninger et al., 2004). Such information could be used to fine-tune existing as well as develop novel approaches to face recognition.

In addition to work on face recognition, considerable attention has been devoted to the

automatic recognition of facial expressions. Early work in this area has focused mainly on recognition of the six prototypical or universal expressions (these are angry, disgust, fear, happy, sad, and surprised; see Ekman et al., 1969) whereas in later work the focus has been to provide a more fine-grained recognition and even interpretation of core *components* of facial expressions.

As mentioned above, all computational systems follow the same basic structure of face detection, feature extraction, and classification. Moreover, one can again divide the proposed systems into different categories on the basis of whether they use holistic information, local feature-based information, or a hybrid approach. An additional aspect is whether the systems estimate the *deformation* of a neutral face for each image or whether they rely explicitly on *motion* to detect a facial expression.

Systems based on *holistic information* have employed PCA (Calder et al., 2001) to recognize static images, estimating dense optic flow to analyze the deformation of the face in two dimensions (Bartlett et al., 1999), as well as 3D deformable face models (DeCarlo and Metaxas, 1996). In contrast, systems based on *local information* rely on analyzing regions in the face that are prone to changes under facial expressions. Such systems have initially used tracking of 2D contours (Terzopoulos and Waters, 1993) or high-contrast regions (Rosenblum et al., 1996; Black and Yacoob, 1997), elastic bunch-graph matching with Gabor filters (Zhang et al., 1998), as well higher-level 3D face models (Gokturk et al., 2002). Despite good recognition results on a few existing test databases, however, these systems mostly focused on recognition of the six prototypical expressions. Therefore, they could not extract important dimensions such as the intensity of the recognized expressions — dimensions, which are critical for human–computer interface (HCI) applications, for example.

A very influential description of facial expressions is FACS developed by Ekman and Friesen in the late 1960s and continuously improved in the following years. As mentioned above (see section on recognition of expressions), FACS encodes both anatomical muscle activations as well as so-called miscellaneous actions in 44 action units. It is important to stress again that the co-activation of action units into complex expressions is external to FACS, making it a purely descriptive rather than an inferential system. In addition, the mapping from action units to complex expressions is ambiguous. Nevertheless, most recent systems try to recognize action units from still images or image sequences — perhaps due to the fact that FACS is one of the few parametric, high-level descriptions of facial motion. As a result of the highly localized structure of action units, these systems rely mostly on local information and use a combination of low-level, appearance-based features and geometric facial components for increased robustness (Bartlett et al., 1999; Donato et al., 1999). One of the most advanced recognition systems (Tian et al., 2001) uses a hybrid scheme combining both low- and high-level local information in a neural network to recognize 16 action units from static images with an accuracy of 96%.

Similarly to face recognition systems there exist several standard databases for benchmarking facial expression algorithms (Ekman and Friesen, 1978; Kanade et al., 2000) — so far, however, no comprehensive benchmark comparing the different systems for facial expression analysis has been developed.

Interestingly, several of the systems discussed here have been explicitly benchmarked against human performance — both in the case of prototypical expressions (Calder et al., 2001) and in the case of action unit recognition (Bartlett et al., 1999; Tian et al., 2001). This shows that in the area of facial expression recognition, the coupling between psychological and computational research is much closer than that seen in identity recognition — one of the reasons may be that expression analysis has drawn more heavily from results in psychological research (Ekman et al., 1969; Ekman and Friesen, 1978).

Finally, the following presents a list of key challenges in the area of automatic recognition of facial expressions that still need to be addressed to design and implement robust systems with human-like performance:

Robust extraction of facial features: Even more than in the case of identity recognition,

extraction of the *exact* shape of facial components would enable to determine action units very precisely.

Dealing with variations in appearance: In contrast to identity recognition, expressions need to be recognized *despite* changes in gender and identity (as well as of course additional parameters such as pose, lighting, etc.). Current algorithms do not yet perform these generalizations well enough.

Going beyond FACS: Although FACS has remained popular, it is still unclear whether it really is a useful basis for describing the space of facial expressions — both from a human perspective as well as from a statistical point of view. Among the alternatives that have been proposed is FACS+, an encoding system for facial actions that seeks to determine well-defined actions as well as the mapping between these actions and complex expressions (Essa and Pentland, 1997). Another alternative is the MPEG4 face-coding scheme, a very generic face animation framework based on movement of keypoints on the face (e.g., Koenen, 2000; see also Section "Dynamic information" above).

Full spatio-temporal, high-level models for recognition: Facial expressions are a highly efficient form of communication — the communicative aspect is not yet exploited by current systems, which is partially due to the lack of explicit models of how humans employ facial motions to convey meaning. Such knowledge could, for example, prove useful as a high-level prior for automatic recognition of expressions.

More descriptive power: For humans, recognition of the expression is only the first step in a longer pipeline, which involves not only judgments of intensity, but also other interpretative dimensions such as believability, naturalness, etc. Our behaviour may be more determined by these dimensions rather than by the classification itself.

Summary and conclusions

The review of psychophysical studies showed that faces are processed in terms of their components and their spatial relationship (configural information). The integrative model by Schwaninger et al. (2002, 2003a) provides a good basis for combining the component-configural hypothesis and holistic aspects of face processing. According to the model, component and configural information are first analyzed separately and then integrated for recognition. Rotating the faces in the plane results in a strong impairment of configural processing, while component processing is much less, if at all, affected by plane rotation. This could be due to capacity limitations of an orientation normalization mechanism such as mental rotation, which is required in order to extract configural information from the plane rotated or inverted faces. Because adult face recognition relies more on the processing of configural information than basic level object recognition, a strong inversion effect is obtained.

Different facial areas and facial motions are important for the recognition of different emotions. Most of the models on facial expression processing have stressed the importance of component information while some models also integrate configural information (e.g., Izard et al., 1983; White, 2000). As pointed out by Schwaninger et al. (2002, 2003a), a model which assumes separate processing of component and configural information before integrating them can explain the effects of facial expression processing in the Thatcher illusion and the face composite illusion. In upright faces, both component and configural information can be processed. This results in a bizarre facial expression in the Thatcher illusion and in a new identity or facial expression when different face composites are aligned. Turning faces upside-down disrupts configural processing. As a consequence, the bizarre facial expression in the Thatcher illusion vanishes. Similarly, the absence of interference from configural information in inverted composites makes it easier to identify the different identities (Young et al., 1987) or emotions (Calder et al., 2000).

The model by Bruce and Young (1986) proposes separate parallel routes for recognizing facial identity, facial expression, and speech. Recent physiological models proposed by Haxby et al. (2000), O'Toole et al. (2002), and Adolphs (2002) are

consistent with this view. Although now much is known about the role and interaction of different brain areas in recognition of identity and expression, the neuronal implementation of analyzing component and configural information and their integration with motion information is less clear.

Interestingly, both identity and expression recognition in computer vision follow the same basic processing steps. Following face detection, in a second step the image area containing a face is processed to extract discriminative features, which are then compared to a database of stored identities or expressions. The different recognition algorithms can be distinguished on whether they use holistic information, local feature-based information, or a hybrid approach; the last two are usually more successful than the first one. One example of a close connection between computational modeling and psychophysical research in this context is the set of studies by Wallraven et al. (2002, 2005b) on the implementation and validation of a model of component and configural processing in identity recognition. On the basis of the model developed by Schwaninger et al. (2003a, 2004) outlined above, they implemented the two routes of face processing using methods from computer vision. The building blocks of the face representation consisted of local features that were extracted at salient image regions at different spatial frequency scales. The basic idea for the implementation of the two routes was that configural processing should be based on a global, position-sensitive connection of these features, whereas component processing should be local and position insensitive. Using these relatively simple ingredients, they showed that the model could capture the human performance pattern observed in the psychophysical experiments. As can be seen in Fig. 4, the results of the computational model are very similar to the psychophysical results obtained with humans in the experiment conducted by Schwaninger et al. (2002) (see also Fig. 3).

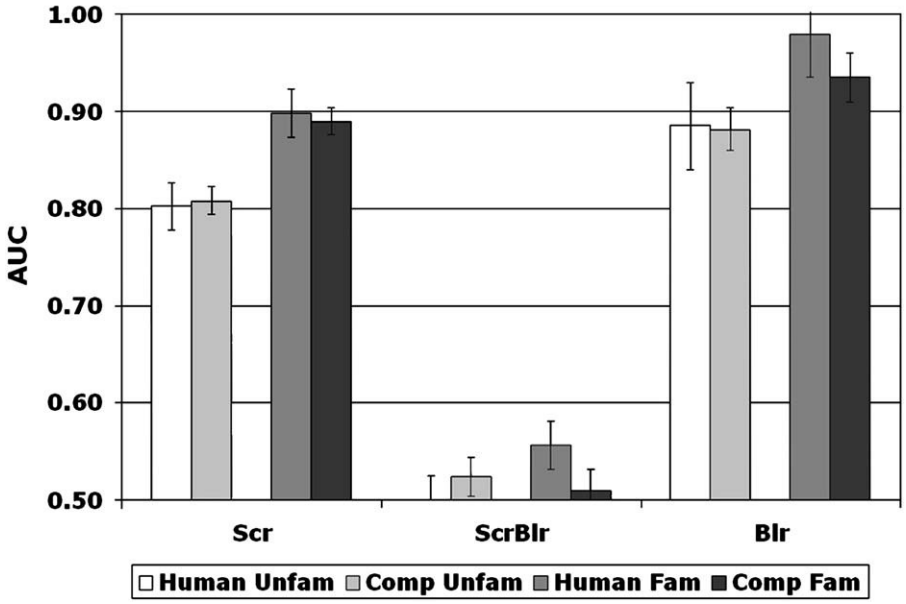

Fig. 4. Human and computational (Comp) performance for the face recognition task of Schwaninger et al. (2002) for unfamiliar (Unfam) and familiar (Fam) face recognition (adapted from Wallraven et al., 2005b). Performance is shown as area under the ROC-curve (AUC). In order to determine the relative contributions of component and configural processing, participants had to recognize previously learnt faces in either scrambled (Scr), blurred (Blr), or scrambled-blurred (ScrBlr) conditions (also see human data in Fig. 3). The recognition performance of the computational model (Comp) was very similar to human performance. Moreover, the results indicate that the two-route implementation of configural and component processing captures the relative contributions of either route to recognition. The observed increase for familiar face recognition by humans could also be modeled with the computational system.

Moreover, combining the two routes resulted in increased recognition performance, which has implications in a computer vision context regarding recognition despite changes in viewpoint (see Wallraven et al., 2005b).

In general, a closer coupling between psychophysical research, neuroscience, and computer vision would benefit all research areas by enabling a more advanced statistical analysis of the information necessary to recognize individuals and expressions as well as the development of better, perceptually motivated recognition algorithms that are able to match human classification performance. This will be necessary in order to better understand processing of component, configural, and motion information and their integration for recognition of identity and facial expression.

Abbreviations

FACS	facial action coding system
FERET	FacE REcognition Technology
FFA	fusiform face area
FLMP	fuzzy logical model of perception
HCI	human–computer interface
ICA	independent component analysis
LDA	linear discriminant analysis
MT	middle temporal
NMF	non-negative matrix factorization
OFA	occipital face area
STS	superior temporal sulcus
SVM	support vector machine

References

Adolphs, R. (2002) Recognizing emotion from facial expressions: psychological and neurological mechanisms. Behav. Cogn. Neurosci. Rev., 1(1): 21–61.

Andrews, T.J. and Ewbank, M.P. (2004) Distinct representations for facial identity and changeable aspects of faces in human temporal lobe. NeuroImage, 23: 905–913.

Bahrick, H.P., Bahrick, P.O. and Wittlinger, R.P. (1975) Fifty years of memory for names and faces: a cross-sectional approach. J. Exp. Psychol.: Gen., 104: 54–75.

Bartlett, M.S., Hager, J.C., Ekman, P. and Sejnowski, T.J. (1999) Measuring facial expressions by computer image analysis. Psychophysiology, 36: 253–263.

Bartlett, M.S., Movellan, J.R. and Sejnowski, T.J. (2002) Face recognition by independent component analysis. IEEE Trans. Neural Networks, 13(6): 1450–1464.

Bassili, J.N. (1978) Facial motion in the perception of faces and of emotional expression. J. Exp. Psychol.: Hum. Percep. Perform., 4(3): 373–379.

Bassili, J. (1979) Emotion recognition: the role of facial motion and the relative importance of upper and lower areas of the face. J. Pers. Soc. Psychol., 37: 2049–2059.

Belhumeur, P., Hespanha, J. and Kriegman, D. (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal.Mach. Intell., 19(7): 711–720.

Bentin, S., Allison, T., Puce, A., Perez, E. and McCarthy, G. (1996) Electrophysiological studies of face perception in human. J. Cogn. Neurosci., 8: 551–565.

Bentin, S., Deouell, L.Y. and Soroker, N. (1999) Selective visual streaming in face recognition: evidence from developmental prosopagnosia. NeuroReport, 10: 823–827.

Biederman, I. (1987) Recognition-by-components: a theory of human image understanding. Psychol. Rev., 94(2): 115–147.

Biederman, I. and Kalocsai, P. (1997) Neurocomputational bases of object and face recognition. Philos. Trans. R. Soc. Lond. Ser. B, 352: 1203–1219.

Black, M. and Yacoob, Y. (1997) Recognizing facial expressions in image sequences using local parameterized models of image motion. Int. J. Comput. Vis., 25(1): 23–48.

Blanz, V. and Vetter, T. (1999) A morphable model for the synthesis of 3D faces. SIGGRAPH'99, Conference Proceedings, Los Angeles, CA, USA, pp. 187–194.

Blanz, V. and Vetter, T. (1993) Face recognition based on fitting a 3D morphable model. IEEE Trans. Pattern Anal. Mach. Intell., 25: 1063–1074.

Bobes, M.A., Martín, M., Olivares, E. and Valdés-Sosa, M. (2000) Different scalp topography of brain potentials related to expression and identity matching of faces. Cogn. Brain Res., 9: 249–260.

Boutet, I., Collin, C. and Faubert, J. (2003) Configural face encoding and special frequency information. Percept. Psychophys., 65(7): 1087–1093.

Bruce, V. (1988) Recognising Faces. Lawrence Erlbaum Associates, Hillsdale, NJ.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P.J.B., Burton, A.M. and Miller, P. (1999) Verification of face identities from images captured on video. J. Exp. Psychol. Appl., 5(4): 339–360.

Bruce, V., Henderson, Z., Newman, C. and Burton, M.A. (2001) Matching identities of familiar and unfamiliar faces caught on CCTV images. J. Exp. Psychol.: Appl., 7: 207–218.

Bruce, V. and Young, A. (1986) Understanding face recognition. Brit. J. Psychol., 77: 305–327.

Brunelli, R. and Poggio, T. (1993) Face recognition: features versus templates. IEEE Trans. Pattern Anal. Mach. Intell., 15(10): 1042–1052.

Bukach, C.M., Gauthier, I. and Tarr, M.J. (in press) Beyond faces and modularity: the power of an expertise framework. Trends Cogn. Sci., 10(4): 159–166.

Calder, A.J., Burton, A.M., Miller, P., Young, A.W. and Akamatsu, S. (2001) A principal component analysis of facial expressions. Vis. Res., 41: 1179–1208.

Calder, A.J. and Young, A.W. (2005) Understanding the recognition of facial identity and facial expression. Nat. Rev. Neurosci., 6: 641–651.

Calder, A.J., Young, A.W., Keane, J. and Deane, M. (2000) Configural information in facial expression perception. J. Exp. Psychol., 26(2): 527–551.

Chao, L.L., Haxby, J.V. and Martin, A. (1999) Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat. Neurosci., 2: 913–919.

Christie, F. and Bruce, V. (1998) The role of dynamic information in the recognition of unfamiliar faces. Mem. Cogn., 26(4): 780–790.

Collishaw, S.M. and Hole, G.J. (2000) Featural and configurational processes in the recognition of faces of different familiarity. Perception, 29: 893–910.

Cootes, T., Edwards, G. and Taylor, C. (2001) Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell., 23: 681–685.

Cunningham, D., Kleiner, M., Wallraven, C. and Bülthoff, H. (2005) Manipulating video sequences to determine the components of conversational facial expressions. ACM Trans. App. Percept., 2(3): 251–269.

Damasio, A.R., Damasio, H. and Van Hoesen, G.W. (1982) Prosopagnosia: anatomic bases and behavioral mechanisms. Neurology, 32: 331–341.

Damasio, A.R., Tranel, D. and Damasio, H. (1990) Face agnosia and the neural substrates of memory. Annu. Rev. Neurosci., 13: 89–109.

Darwin, C. (1872) The Expression of the Emotions in Man and Animals. John Murray, London.

Davidoff, J. and Donnelly, N. (1990) Object superiority: a comparison of complete and part probes. Acta Psychol., 73: 225–243.

Davidoff, J. and Landis, T. (1990) Recognition of unfamiliar faces in prosopagnosia. Neuropsychologia, 28: 1143–1161.

DeCarlo, D. and Metaxas, D., 1996. The integration of optical flow and deformable models with applications to human face shape and motion estimation. Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR '96), San Francisco, CA, USA, pp. 231–238.

DeGelder, B., Frissen, I., Barton, J. and Hadjikhani, N. (2003) A modulatory role for facial expressions in prosopagnosia. Proc. Natl. Acad. Sci. USA, 100(22): 13105–13110.

DeRenzi, E. (1997) Prosopagnosia. In: Feinberg, T.E. and Farah, M.J. (Eds.), Behavioral Neurology and Neuropsychology. McGraw-Hill, New York, pp. 245–256.

Diamond, R. and Carey, S. (1986) Why faces are and are not special: an effect of expertise. J. Exp. Psychol.: Gen., 115(2): 107–117.

Donato, G., Bartlett, S., Hager, J., Ekman, P. and Sejnowski, T. (1999) Classifying facial actions. IEEE Trans. Pattern Anal. Mach. Intell., 21(10): 974–989.

Downing, P.E., Chan, A.W., Peelen, M.V., Dodds, C.M. and Kanwisher, N. (2005). Domain specificity in visual cortex. Cereb. Cortex (Dec 7, electronic publication, ahead of print).

Duchenne, B. (1990) The Mechanism of Human Facial Expression or an Electro-Physiological Analysis of the Expression of the Emotions. Cambridge University Press, New York.

Edwards, K. (1998) The face of time: temporal cues in facial expressions of emotion. Psychol. Sci., 9: 270–276.

Eimer, M. and Holmes, A. (2002) An ERP study on the time course of emotional face processing. Cogn. Neurosci. Neuropsychol., 13(4): 427–431.

Ekman, P. and Friesen, W.F. (1978) Facial Action Coding System. Consulting Psychologists Press, Palo Alto.

Ekman, P. and Friesen, W.V. (1982) Felt, false, and miserable smiles. J. Nonverb. Behav., 6: 238–252.

Ekman, P., Hager, J., Methvin, C. and Irwin, W. (1969) Ekman-Hager Facial Action Exemplars. Human Interaction Laboratory, University of California, San Francisco.

Ellison, J.W. and Massaro, D.W. (1997) Featural evaluation, integration, and judgment of facial affect. J. Exp. Psychol.: Hum. Percept. Perform., 23(1): 213–226.

Essa, I. and Pentland, A. (1994) A vision system for observing and extracting facial action parameters. Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR'94), Seattle, WA, USA, pp. 76–83.

Essa, I. and Pentland, A. (1997) Coding, analysis, interpretation and recognition of facial expressions. IEEE Trans. Pattern Anal. Mach. Intell., 19: 757–763.

Farah, M.J., Levinson, K.L. and Klein, K.L. (1995a) Face perception and within-category discrimination in prosopagnosia. Neuropsychologia, 33: 661–674.

Farah, M.J., Tanaka, J.W. and Drain, H.M. (1995b) What causes the face inversion effect? J. Exp. Psychol.: Hum. Percept. Perform., 21(3): 628–634.

Frijda, N.H. and Philipszoon, E. (1963) Dimensions of recognition of emotion. J. Abnorm. Soc. Psychol., 66: 45–51.

Frois-Wittmann, J. (1930) The judgment of facial expression. J. Exp. Psychol., 13: 113–151.

Gauthier, I., Behrmann, M. and Tarr, M.J. (1999a) Can face recognition be dissociated from object recognition? J. Cogn. Neurosci., 11: 349–370.

Gauthier, I., Skudlarski, P., Gore, J.C. and Anderson, A.W. (2000a) Expertise for cars and birds recruits brain areas involved in face recognition. Nat. Neurosci., 3: 191–197.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P. and Gore, J.C. (1999b) Activation of the middle fusiform area increases with expertise in recognizing novel objects. Nat. Neurosci., 6: 568–573.

Gauthier, I., Tarr, M.J., Moylan, J., Skudlarski, P., Gore, J.C. and Anderson, A.W. (2000b) The fusiform "face area" is part of a network that processes faces at the individual level. J. Cogn. Neurosci., 12(3): 495–504.

Gibson, J.J. (1957) Optical motions and transformations as stimuli for visual perception. Psychol. Rev., 64: 228–295.

Gibson, J.J. (1966) The Senses Considered as Perceptual Systems. Houghton Mifflin, Boston, MA.

Gokturk, S., Tomasi, C., Girod, B. and Bouguet, J. (2002) Model-based face tracking for view-independent facial expression recognition. In: Fifth IEEE International Conference

on Automatic Face and Gesture Recognition, Washington, D.C., USA, pp. 287–293.
Hanawalt, N. (1944) The role of the upper and lower parts of the face as the basis for judging facial expressions: II. In posed expressions and "candid camera" pictures. J. Gen. Psychol., 31: 23–36.
Hasselmo, M.E., Rolls, E.T. and 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. Exp. Brain Res., 32: 203–218.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L. and Pietrini, P. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science, 293: 2425–2430.
Haxby, J.V., Hoffman, E.A. and Gobbini, M.I. (2000) The distributed human neural system for face perception. Trends Cogn. Sci., 4(6): 223–233.
Heisele, B., Ho, P., Wu, J. and Poggio, T. (2003) Face recognition: comparing component-based and global approaches. Comput. Vis. Image Understanding, 91(1–2): 6–21.
Henke, K., Schweinberger, S.R., Grigo, A., Klos, T. and Sommer, W. (1998) Specificity of face recognition: recognition of exemplars of non-face objects in prosopagnosia. Cortex, 34(2): 289–296.
Hill, H. and Johnson, A. (2001) Categorization and identity from the biological motion of faces. Curr. Biol., 11: 880–885.
Hjelmas, E. and Low, B. (2001) Face detection: a survey. Comput. Vis. Image Understanding, 83: 236–274.
Hoffman, E. and Haxby, J. (2000) Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat. Neurosci., 3: 80–84.
Humphreys, G., Donnelly, N. and Riddoch, M. (1993) Expression is computed separately from facial identity, and is computed separately for moving and static faces: neuropsychological evidence. Neuropsychologia, 31: 173–181.
Ishai, A., Haxby, J.V. and Ungerleider, L.G. (2002) Visual imagery of famous faces: effects of memory and attention revealed by fMRI. NeuroImage, 17: 1729–1741.
Ishai, A., Ungerleider, L.G., Martin, A. and Haxby, J.V. (2000) The representation of objects in the human occipital and temporal cortex. J. Cogn. Neurosci., 12: 35–51.
Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L. and Haxby, J.V. (1999) Distributed representation of objects in the human ventral visual pathway. Proc. Natl. Acad. Sci. USA, 96: 9379–9384.
Izard, C.E. (1979) The maximally discriminative facial movement coding system (MAX). Unpublished manuscript. (Available from Instructional Resource Center, University of Delaware, Newark, DE.)
Izard, C.E., Dougherty, L.M. and Hembree, E.A. (1983) A system for identifying affect expressions by holistic judgments. Unpublished manuscript, University of Delaware.
Johonnson, G. (1973) Visual perception of biological motion and a model for its analysis. Percept. Psychophys., 14: 201–211.
Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S. and Akamatsu, S. (2001) Dynamic properties influence the perception of facial expressions. Perception, 30: 875–887.
Kanade, T. (1973) Computer Recognition of Human Faces. Basel and Stuttgart, Birkhauser.
Kanade, T., Cohn, J.F. and Tian, Y. (2000) Comprehensive database for facial expression analysis. Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition, Grenoble, France, pp. 46–53.
Kanwisher, N., McDermott, J. and Chun, M.M. (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci., 17: 4302–4311.
Kanwisher, N. and Yovel, G. (in press) The fusiform face area: a cortical region specialized for the perception of faces. Philos. Trans. R. Soc. Lond. Ser. B.
Knappmeyer, B., Thornton, I.M. and Bülthoff, H.H. (2003) The use of facial motion and facial form during the processing of identity. Vis. Res., 43: 1921–1936.
Koenen, R. (2000) Mpeg-4 Project Overview, International Organization for Standardization, ISO/IEC JTC1/SC29/WG11.
Köhler, W. (1940) Dynamics in Psychology. Liveright, New York.
Lander, K. and Bruce, V. (2000) Recognizing famous faces: exploring the benefits of facial motion. Ecol. Psychol., 12(4): 259–272.
Lander, K., Bruce, V. and Hill, H. (2001) Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. Appl. Cogn. Psychol., 15: 101–116.
Lander, K., Christie, F. and Bruce, V. (1999) The role of movement in the recognition of famous faces. Mem. Cogn., 27(6): 974–985.
Leder, H. and Bruce, V. (1998) Local and relational aspects of face distinctiveness. Quart. J. Exp. Psychol., 51A(3): 449–473.
Leder, H. and Bruce, V. (2000) When inverted faces are recognized: the role of configural information in face recognition. Quart. J. Exp. Psychol., 53A(2): 513–536.
Leder, H., Candrian, G., Huber, O. and Bruce, V. (2001) Configural features in the context of upright and inverted faces. Perception, 30: 73–83.
Lee, D. and Seung, H. (1999) Learning the parts of objects by non-negative matrix factorization. Nature, 401: 788–791.
Leopold, D.A., O'Toole, A., Vetter, T. and Blanz, V. (2001) Prototype-referenced shape encoding revealed by high-level after effects. Nat. Neurosci., 4: 89–94.
Leventhal, H. and Sharp, E. (1965) Facial expression as indicators of distress. In: Tomkins, S.S. and Izard, C.E. (Eds.), Affect, Cognition and Personality: empirical Studies. Springer, New York, pp. 296–318.
Lewis, M.B. and Edmonds, A.J. (2003) Face detection: mapping human performance. Perception, 32(8): 903–920.
Li, B. and Chellappa, R. (2001) Face verification through tracking facial features. J. Op. Soc. Am. A, 18(12): 2969–2981.
Li, S. and Jain, A. (Eds.). (2004) Handbook of Face Recognition. Springer, New York.
Liu, J., Harris, A. and Kanwisher, N. (2002) Stages of processing in face perception: an MEG study. Nat. Neurosci., 5: 910–916.
Marr, D. (1982) Vision. Freeman Publishers, San Francisco.
McNeil, J.E. and Warrington, E.K. (1993) Prosopognosia: a face-specific disorder. Quart. J. Exp. Psychol., 46A: 1–10.

Moscovitch, M., Winocur, G. and Behrmann, M. (1997) What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. J. Cogn. Neurosci., 9: 555–604.

Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T. and Vatikiotis-Bateson, E. (2004) Visual prosody and speech intelligibility: head movement improves auditory speech perception. Psychol. Sci., 15(2): 133–137.

Murray, J.E., Yong, E. and Rhodes, G. (2000) Revisiting the perception of upside-down faces. Psychol. Sci., 11: 498–502.

Nummenmma, T. (1964) The language of the face. In: Jyvaskyla Studies in Education, Psychology, and Social Research. Jyvaskyla, Finland.

O'Craven, K.M., Downing, P.E. and Kanwisher, N. (1999) FMRI evidence for objects as the units of attentional selection. Nature, 401: 584–587.

O'Toole, A.J., Phillips, P.J., Cheng, Y., Ross, B. and Wild, H.A. (2000) Face recognition algorithms as models of human face processing. Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France.

O'Toole, A.J., Roark, D.A. and Abdi, H. (2002) Recognizing moving faces: a psychological and neural synthesis. Trends Cogn. Sci., 6(6): 261–266.

Perret, D.I., Hietanen, J.K., Oram, M.W. and Benson, P.J. (1992) Organization and functions of cells in the macaque temporal cortex. Philos. Trans. R. Soc. Lond. Ser. B, 335: 23–50.

Perret, D.I., Mistlin, A.J. and Chitty, A.J. (1987) Visual neurones responsive to faces. Trends Neurosci., 10: 358–364.

Perret, D.I. and Oram, M.W. (1993) Image Vis. Comput., 11: 317–333.

Perret, D.I., Rolls, E.T. and Caan, W. (1982) Visual neurons responsive to faces in the monkey temporal cortex. Exp. Brain Res., 47: 329–342.

Phillips, P.J. (1999) Support vector machines applied to face recognition. Adv. Neural Inform. Process. Systems, 11: 803–809.

Pike, G., Kemp, R., Towell, N. and Phillips, K. (1997) Recognizing moving faces: the relative contribution of motion and perspective view information. Vis. Cogn., 4: 409–437.

Plutchik, R. (1962) The Emotions: Facts, Theories, and A New Model. Random House, New York.

Puce, A., Allison, T., Bentin, S., Gore, J.C. and McCarthy, G. (1998) Temporal cortex activation in humans viewing eye and mouth movements. J. Neurosci., 18: 2188–2199.

Rhodes, G., Brake, S. and Atkinson, A.P. (1993) What's lost in inverted faces? Cognition, 47: 25–57.

Roack, D.A., Barrett, S.E., Spence, M., Abdi, H. and O'Toole, A.J. (2003) Memory for moving faces: psychological and neural perspectives on the role of motion in face recognition. Behav. Cogn. Neurosci. Rev., 2(1): 15–46.

Rock, I. (1973) Orientation and Form. Academic Press, New York.

Rock, I. (1974) The perception of disoriented figures. Sci. Am., 230: 78–85.

Rock, I. (1988) On Thompson's inverted-face phenomenon (Research Note). Perception, 17: 815–817.

Rosenblum, M., Yacoob, Y. and Davis, L. (1996) Human expression recognition from motion using a radial basis function network architecture. IEEE Trans. Neural Networks, 7(5): 1121–1138.

Rossion, B., Caldara, R., Seghier, M., Schuller, A.M., Lazeyras, F. and Mayer, E. (2003) A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. Brain, 126: 2381–2395.

Rossion, B., Curran, T. and Gauthier, I. (2002) A defense of the subordinate-level expertise account for the N170 component. Cognition, 85: 189–196.

Rotshtein, P., Henson, R.N.A., Treves, A., Driver, J. and Donlan, R.J. (2005) Morphing Marilyn into Maggie dissociates physical identity face representations in the brain. Nat. Neurosci., 8(1): 107–113.

Sayette, M.A., Cohn, J.F., Wertz, J.M., Perrott, M.A. and Dominic, J. (2001) A psychometric evaluation of the facial action coding system for assessing spontaneous expression. J. Nonverb. Behav., 25: 167–186.

Schwaninger, A., Carbon, C.C. and Leder, H. (2003a) Expert face processing: specialization and constraints. In: Schwarzer, G. and Leder, H. (Eds.), Development of Face Processing. Göttingen, Hogrefe, pp. 81–97.

Schwaninger, A., Lobmaier, J. and Collishaw, S.M. (2002) Component and configural information in face recognition (Lectures Notes). Comput. Sci., 2525: 643–650.

Schwaninger, A., Lobmaier, J. and Fischer, M. (2005) The inversion effect on gaze perception is due to component information. Exp. Brain Res., 167: 49–55.

Schwaninger, A. and Mast, F. (2005) The face inversion effect can be explained by capacity limitations of an orientation normalization mechanism. Jpn. Psychol. Res., 47(3): 216–222.

Schwaninger, A., Ryf, S. and Hofer, F. (2003b) Configural information is processed differently in perception and recognition of faces. Vis. Res., 43: 1501–1505.

Schwaninger, A., Wallraven, W. and Bülthoff, H.H. (2004) Computational modeling of face recognition based on psychophysical experiments. Swiss J. Psychol., 63(3): 207–215.

Searcy, J.H. and Bartlett, J.C. (1996) Inversion and processing of component and spacial-relational information in faces. J. Exp. Psychol.: Hum. Percept. Perform., 22(4): 904–915.

Sekuler, A.B., Gaspar, C.M., Gold, J.M. and Bennett, P.J. (2004) Inversion leads to quantitative, not qualitative, changes in face processing. Curr. Biol., 14(5): 391–396.

Sergent, J. (1984) An investigation into component and configural processes underlying face perception. Br. J. Psychol., 75: 221–242.

Sergent, J. (1985) Influence of task and input factors on hemispheric involvement in face processing. J. Exp. Psychol.: Hum. Percept. Perform., 11(6): 846–861.

Swets, D. and Weng, J. (1996) Using discriminant eigenfeatures for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell., 18: 831–836.

Tanaka, J.W. and Farah, M.J. (1991) Second-order relational properties and the inversion effect: testing a theory of face perception. Percept. Psychophys., 50(4): 367–372.

Tanaka, J.W. and Farah, M.J. (1993) Parts and wholes in face recognition. Quart. J. Exp. Psychol., 46A(2): 225–245.

Terzopoulos, D. and Waters, K. (1993) Analysis and synthesis of facial image sequences using physical and anatomical models. IEEE Trans. Pattern Anal. Mach. Intell., 15: 569–579.

Thompson, P. (1980) Margaret Thatcher: a new illusion. Perception, 9: 483–484.

Tian, Y., Kanade, T. and Cohn, J. (2001) Recognizing action units for facial expression analysis. IEEE Trans. Pattern Anal. Mach. Intell., 23(2): 97–115.

Tronick, E., Als, H. and Brazelton, T.B. (1980) Monadic phases: a structural descriptive analysis of infant-mother face-to-face interaction. Merrill-Palmer Quart. Behav. Dev., 26: 3–24.

Turk, M. and Pentland, A. (1991) Eigenfaces for recognition. J. Cogn. Neurosci., 3: 72–86.

Valentine, T. (1995) Cognitive and Computational Aspects of Face Recognition: Explorations in Face Space. Routledge, London.

Valentine, T. and Bruce, V. (1988) Mental rotation of faces. Mem. Cogn., 16(6): 556–566.

Vuilleumier, P., Armony, J.L., Driver, J. and Dolan, R.J. (2001) Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron, 30: 829–841.

Wacholtz, E. (1996) Can we learn from the clinically significant face processing deficits, prosopagnosia and capgras delusion? Neuropsychol. Rev., 6: 203–258.

Wachsmuth, E., Oram, M.W. and Perret, D.I. (1994) Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. Cereb. Cortex, 4: 509–522.

Wallraven, C., Breidt, M., Cunningham, D.W. and Bülthoff, H.H. (2005a) Psychophysical evaluation of animated facial expressions. Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization, A Coruña, Spain, pp. 17–24.

Wallraven, C., Schwaninger, A. and Bülthoff, H.H. (2005b) Learning from humans: computational modeling of face recognition. Network: Comput. Neural Syst, 16(4): 401–418.

Wallraven, C., Schwaninger, A., Schuhmacher, S. and Bülthoff, H.H. (2002) View-based recognition of faces in man and machine: re-visiting inter-extra-ortho (Lectures Notes). Comput. Sci., 2525: 651–660.

White, M. (2000) Parts and wholes in expression recognition. Cogn. Emotion, 14(1): 39–60.

Williams, M.A., Moss, S.A. and Bradshaw, J.L. (2004) A unique look at face processing: the impact of masked faces on the processing of facial features. Cognition, 91: 155–172.

Winston, J.S., Henson, R.N.A., Fine-Goulden, M.R. and Dolan, R.J. (2004) fMRI-adaption reveals dissociable neural representations if identity and expression in face perception. J. Neurophysiol., 92: 1830–1839.

Wiskott, L., Fellous, J. and von der Malsburg, C. (1997) Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Anal. Mach. Intell., 19: 775–779.

Xu, Y. (2005) Revisiting the role of the fusiform face area in visual expertise. Cereb. Cortex, 15(8): 1234–1242.

Xu, Y., Liu, J. and Kanwisher, N. (2005) The M170 is selective for faces, not for expertise. Neuropsychology, 43: 588–597.

Yamane, S., Kaji, S. and Kawano, K. (1988) What facial features activate face neurons in the inferotemporal cortex of the monkey? Exp. Brain Res., 73: 209–214.

Yin, R. (1969) Looking at upside-down faces. J. Exp. Psychol., 81(1): 141–145.

Young, A.W., Hellawell, D. and Hay, D.C. (1987) Configural information in face perception. Perception, 16: 747–759.

Yovel, G. and Kanwisher, N. (2005) The neural basis of the behavioral face-inversion effect. Curr. Biol., 15: 2256–2262.

Zhang, Z., Lyons, M., Schuster, M. and Akamatsu, S. (1998) Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron. Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, pp. 454–459.