

Chapter 22

Dynamic Aspects of Face Processing in Humans

Heinrich H. Bülthoff, Douglas W. Cunningham, and Christian Wallraven

22.1 Introduction

The human face is capable of a wide variety of facial expressions that manifest themselves as usually highly non-rigid deformations of the face. On the one hand, this presents the visual system with a problem: Recognizing someone requires determining what information in the face remains constant despite the various facial deformations. Extraction of such invariant features will allow me, for example, to identify my neighbor regardless of whether he or she is smiling or looking sad. On the other hand, the impressive repertoire of changes can also be seen as a positive: It provides considerable information. The particular way my neighbor smiles or looks sad might well be used for identification, similar to how Jack Nicholson's and Tom Cruise's smiles are very specific to them.

In addition to potentially providing information about who someone is, facial deformations can help us to infer something about a person's age, social status, general health, level of fatigue, and focus of attention. Likewise, changes in the facial surface play a central, albeit often ignored, role in communication. Facial deformations that serve this latter role are generally referred to as facial expressions.

A distinction should be drawn between the information that *is* present in a *specific image* and the information that *must be* present for that expression or person to be

H.H. Bülthoff (✉) · D.W. Cunningham · C. Wallraven
Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany
e-mail: heinrich.buelthoff@tuebingen.mpg.de

H.H. Bülthoff · C. Wallraven
Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea

C. Wallraven
e-mail: wallraven@korea.ac.kr

D.W. Cunningham
Brandenburg Technical University, 03046 Cottbus, Germany
e-mail: douglas.cunningham@tu-cottbus.de

recognized. Trying to determine what information is perceptually necessary not only provides critical insights into how humans process faces, but can also yield clues for the design of automated facial recognition and synthesis systems.

Almost all research on the perception of faces—both for identity and expression—has tended to focus on the relatively stable aspects of faces. Some of this information is invariant to deformations of the facial surface, such as the color of or distance between the eyes. In other cases, the result of the deformation *is* the information, such as the shape of the mouth. In such cases, usually the maximum deformation (or peak expression) is examined. In other words, there is a pervasive emphasis on *static* facial information. To some degree, this focus on static information is driven by technology. It is difficult to systematically manipulate a photograph in order to provide the systematic and parameterized variations needed for perceptual experiments without making the photograph look unrealistic. It is considerably more difficult to perform the same manipulations on a *sequence* of images without introducing artifacts either in any given image or across images [29, 80].

In general, however, human faces are not static entities. Indeed, if we meet someone who never moved their face, we would most likely be rather uncomfortable. Some have gone so far as to argue that an individual (specifically, an android) that looks like a human but moves either incorrectly or not at all (i.e., has a “dead” face) will—as a result of the zombie-like appearance—lead to humans being repulsed by that individual [75]. This hypothesis is referred to as the “uncanny valley”. Regardless of whether a zombie—or zombie-like individual—will repulse humans or not, it is clear that the pattern of change over time is itself a great source of information for many visual processes [46] and can often be used to discern real from synthesized stimuli [106].

Fortunately, recent advances in technology, have allowed researchers to carefully and systematically alter video sequences without introducing noticeable artifacts and thus begin to examine the role of motion in face processing. Before one can determine what types of motion are used (i.e., uncover the dynamic features), one must determine if motion plays any role at all. It has been shown that facial motion can provide information about gender [15, 51] and age [14].

In this chapter, we will focus on the role of motion in identity (Sect. 22.2) and expression (Sect. 22.3) recognition in humans, and explain its developmental and neurophysiological aspects. We will make some inferences and conclusions based on results from literature.

22.2 Dynamic Information for Identity

Correctly identifying other people is critical to everyday survival and there has been a considerable amount of research on how such a task might be performed. The literature on how humans use faces to recognize people is quite extensive (for a review, see Chap. 26, this volume or [91]). While the great majority of this literature has focused on static information for identity, there has been an increasing interest in dynamic information (see, e.g., [81]).

Motion can be subdivided into rigid and nonrigid types. Rigid “facial” motion generally refers to the rotations and/or translations of the entire head (e.g., such as nodding or shaking the head). Nonrigid facial motion, in contrast, generally refers to the nonlinear deformations of the facial surface (e.g., lip motion, eyebrow motion). One of the first studies to examine motion presented a 10 second clip of an individual rotating in a chair through a full 360 degrees (rigid motion; [83]). This motion was chosen as it represents a simple change in relative viewpoint, and thus may help an observer to build up a more complete 3D representation of the individual. Accordingly, [83] found higher identity recognition performance in dynamic conditions than in static conditions.

Shortly thereafter, [24] presented contrasting results. They showed five frames of a person moving his/her head up and down and found *no* difference between static and dynamic conditions. The head motion here, they suggested, represents social communication (such as a nod of agreement). Thus the difference in results might be represent a contrast between viewpoint change and social signal. It is important to note that there are a number of other differences between the two studies, such as length of the stimuli, direction of motion (horizontal versus vertical rotations), and task. Subsequent studies have examined some of the differences to determine the source of the conflict.

Another way of saying that [83]’s videos were longer than [24]’s is to say that they contained many more images. This raises the possibility that not only is the difference in results due to the length of the stimuli, but perhaps [83]’s dynamic advantage itself is due merely to the number of images: The dynamic sequence has more images than the static image. Perhaps one of the many views shown in [83]’s had a facial view which was more optimal than the one used in the static condition or in [24]’s five image video. To explicitly test this hypothesis, [69] asked participants to identify a number of famous faces, which were presented in three different formats: as a nine-frame video sequence, a static 3×3 array of the nine images in order, and a static 3×3 array with the nine images in random order. Not only was there a dynamic advantage (despite the sequences being only 9 frames long), but performance 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. It is important to notice, however, that [69]’s sequences consisted primarily of nonrigid motion (talking, expressions) with little rigid motion.

Note that the 9 frames in [69]’s videos were presented sequentially in the dynamic condition and next to one another in the static conditions. Perhaps the mere presence of multiple images is sufficient, but they need to be presented one after another. That is, maybe the images need to be temporally separated or presented at the same spot on the monitor, and motion per se is not required. To test this, [67] and [83] presented a video where the images in the video were presented in a random order. Note that such sequences have motion, but this motion is jerky and erratic (i.e., the motion does not occur in nature). They found that identity was more accurately recognized in normal sequences than in random sequences. This suggests that it is not just the presence of time or motion that is important, but the specific, naturally occurring motion (either horizontal rotation or nonrigid motion) that provides the advantage. As a final, more stringent test that it is the *characteristic motion*

that is important, [67] showed 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. Likewise, they showed that changing the speed of the motion sequence (e.g., by playing parts or all of a video sequence too fast or too slow) decreased recognition performance.

The finding that a scrambled version of simple, uniform, horizontal head rotation was not better than a static photograph, while an intact rotation sequence yielded a significant recognition advantage [83] suggests that something other than characteristic motion might be important. Wallis and Bühlhoff [104], for example, suggested that the temporal coherence of the stimuli provides some information. Specifically, they examined how people learned new faces. While the head of the person-to-be-learned was rotated (horizontally), the identity was changed. One unfamiliar face was shown when the head was at its left extreme position and a different (but still unfamiliar) face when the head was at the right extreme position. The intermediate positions were a morph between the two identities. The results clearly show that participants treated the different identities as if they were the same person (seen from different views). Moreover, the fusion of the two identities was only found for continuous head rotations; Scrambling the order in which the views were presented eliminated the effect. Thus, it seems that spatiotemporal continuity plays a role in learning identity.

In an attempt to determine the specific roles of rigid and nonrigid motion— independent of static information—[52] used motion capture recordings of a conversation to animate an average face. They artificially separated rigid from nonrigid motion and examined identity recognition and sex recognition. They found that both types of motion can be used for both tasks, but that rigid head motion was slightly more useful than nonrigid for identity recognition (while the reverse was true for sex recognition). They also showed that inverting the face and playing the sequence backwards reduced recognition, again pointing to some characteristic motion information.

Following this, [68] also compared the role of different forms of motion. In contrast to [52]’s use of an average face, they used degraded images of familiar individuals (individuals from the same working environment). They found a dynamic advantage for non-rigid motions such as expressions and speech, but not for rigid motions (head nodding). Interestingly, the dynamic advantage was stronger for “distinctive” facial motions. This finding has been extended by [71], who showed that recognition of familiar faces was better when the smile was “natural” rather than “artificial”. Based on these findings, [71] concluded that some familiar faces have characteristic motions that can help in identification via incorporating supplemental motion-based information about the face.

The role of nonrigid motion in the learning of new individuals was examined by [99]. While the novel faces were being learned, the beginning of a smile or a frown was presented (specifically, the first 18 frames). Using a sequential matching paradigm, they showed an advantage of motion when the test image was the same person with a different (static) expression as well as when the same person was seen from a different view (i.e., generalization across expression and viewpoint). The

effect of dynamic information in generalization was subsequently replicated by [84] using the motions from surprise and anger and a delayed visual search paradigm.

Many of the successful demonstrations of a dynamic advantage used degraded stimuli. For example, [65] presented photographic negatives of the faces (see Fig. 22.1b). Likewise, Lander and colleagues have impaired static information in a number of ways, including Gaussian blurring (see Fig. 22.1c), inverting (see Fig. 22.1d), pixelation (see Fig. 22.1e), and thresholding (reducing the image to a two-tone, black/white image; see Fig. 22.1f) [68–70]. As a result, it has been suggested that dynamic information only plays a role when static information is impaired and the person is familiar (see, e.g., [81]). The successful demonstration of a dynamic advantage in nondegraded stimuli of unfamiliar individuals by [52, 84, 99, 104] makes it unlikely that such an explanation captures the whole story.

As an alternate explanation, Thornton and colleagues [99] suggested that previous failures to find a dynamic effect (specifically that of [24]) may have been due to the task used (an old-new task). That is, it might be a memory effect, but not a face perception effect. Additional evidence for this comes from [88], who also used an old-new task to examine the role of different types of motion in learning new individuals. The videos were from a speech database, so contained rigid as well as nonrigid motion. In contrast to the other work on learning new individuals, they found no dynamic advantage, consistent with Thornton's suggestion that the task may be problematic. There is, however, some difficulty with such an explanation: [68] found head nodding did not lead to a dynamic advantage, but they used a naming task (and not an old-new task) for familiar individuals. Interestingly, the type of rigid motion in [68]'s and [24]'s experiments was the same: vertical rotations. It is possible that this also plays a role.

In the first study to explicitly examine the interaction between static and dynamic identity information, [64] recorded several individuals performing various actions (such as chewing and smiling) and used those motions to animate unfamiliar faces. The motions and the faces were subsequently systematically combined. Participants learned, for example, two different faces each with its own motion. Subsequently, they were presented with a series of trials where the motion was always from one individual, but the face shape was a morph of the two. By systematically varying the degree of the morph, they were able to measure psychometric functions showing the independent contribution of shape and motion. They found clear evidence that the characteristic aspects of the motion influenced the learning and subsequent recognition of identity.

22.2.1 Developmental Aspects

As part of the large body of literature on the importance of motion and moving stimuli in general for the perceptual development of the infant, several studies have investigated how infants might benefit from the information inherent in moving faces. Looking at the performance of infants in the context of dynamic face recognition

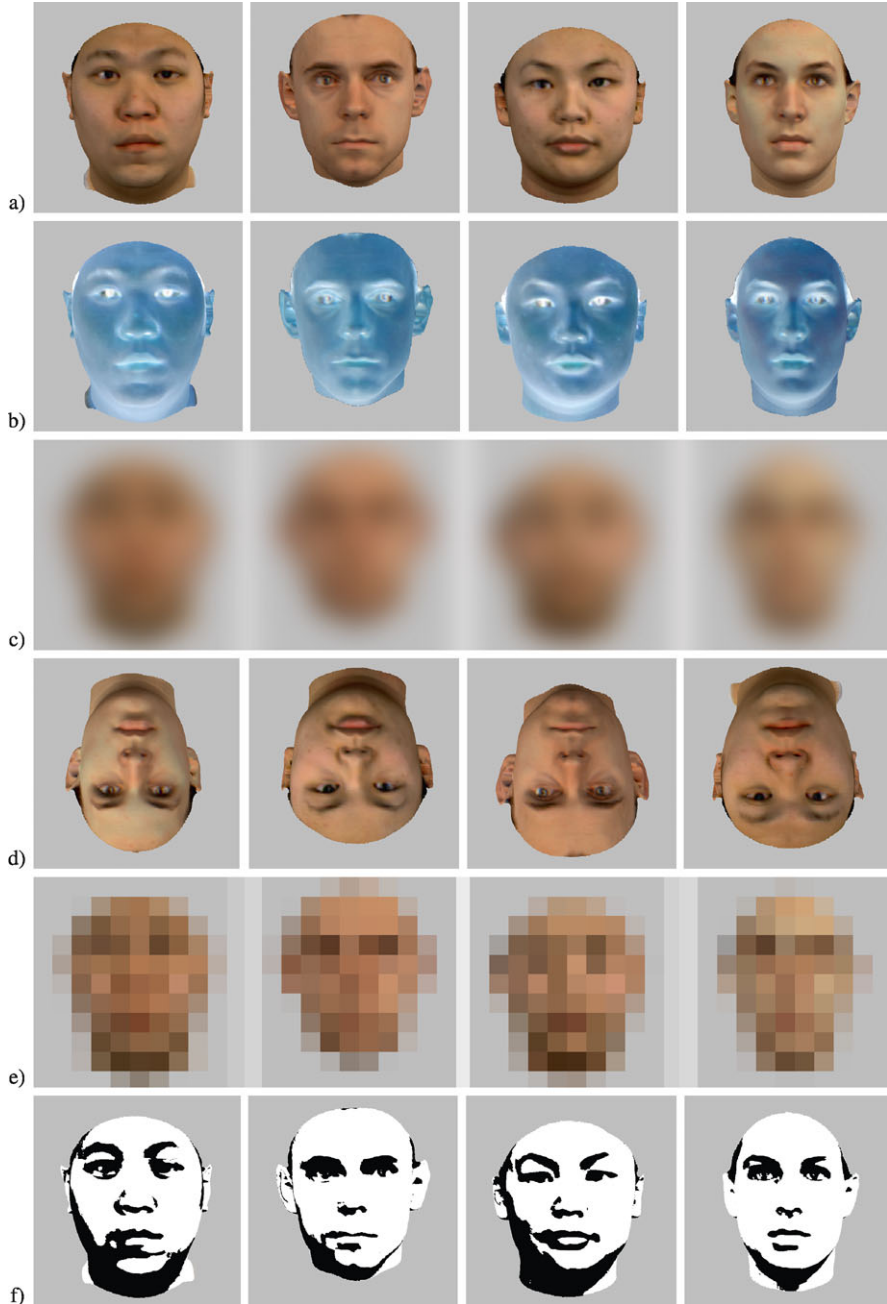


Fig. 22.1 Examples for several individuals of some image degradation procedures **a** the original photos; **b** Photographic negative version of those individuals; **c** Gaussian blur; **d** inversion; **e** pixelation; and **f** thresholding

is especially illuminating for two reasons: First, infants (especially in their first few years) have not yet become full experts in face processing. Indeed, as research has shown, face recognition takes an astonishingly long period to reach adult recognition levels [92]. This means, that infants have not yet reached “ceiling”, that is, they do not yet exhibit the excellent generalization capabilities and almost perfect recognition results usually found in adult observers. A second important reason for investigating infants is that during the first few years, information about facial identity and facial expressions comes exclusively from real, living, moving faces (for example, the gaze of infants is attracted to a movie of the mother’s face versus an abstract movie [55] at as early as 10 weeks of age), whereas for adults the processing of static faces (e.g., photographs) is a much more common activity.

One of the first studies to explicitly test whether infants are specifically sensitive to dynamic information in faces [96] is a follow-up study to [51] using the same stimuli. Since they tested young infants between 4 and 8 months of age, standard recognition paradigms could not be used. Instead, a variant on the preferential looking paradigm was employed. First, an average face told a joke. Whereas the shape of the average face was independent of the actor who told the joke, its deformation was driven by that particular actor’s motion. After having seen such a dynamic face, the infants were presented with the same average face telling a different joke. This time, however, there were two faces: one driven by the previous actor’s motion and another driven by another actor’s motion. The study found that infants tended to spend more time looking at the sequence with the previous actor’s motion indicating that, indeed, the motion signature of the actor was processed—and can be used by infants to disambiguate individuals.

Similarly, in [82] infants between 3 and 4 months of age were tested. In one experiment, infants were either presented with either a short clip of a face in motion (performing a facial expression) or of the last, static frame of that movie. After 30 seconds of exposure, infants then looked at two (static) faces differing in facial expression. The study found that infants could only tell the two faces apart using the dynamic familiarization stage. Only after the familiarization phase was extended to 90 seconds could infants use the static familiarization face for subsequent recognition.

In another related study [94], two groups of 7 and 10 year-old children were required to learn unfamiliar faces from either static or dynamic stimuli with a subsequent static or dynamic test phase. Interestingly, the study found that motion helped during familiarization but not during recognition, that is, it did not matter whether faces during testing were shown as pictures or as movies. On average—and in agreement with the general trend observed during development—older children performed better at the task.

Taken together, these studies show that motion plays perhaps the greatest role during the early stages of development, helping the perceptual system to recognize idiosyncratic movements quickly and enabling a better structural description of faces through the use of motion cues. During adulthood, however, the face processing system has reached an expert level, making it hard to identify the advantage of moving faces for recognition of identity.

22.2.2 *Neurophysiological Aspects*

Given that faces are learned in a dynamic context, one would expect the brain to have specific networks devoted to processing of spatiotemporal information about a face. Indeed, such a hypothesis would go hand in hand with the models proposed for face recognition [17, 20, 50] which posit a separate processing stream for “changeable aspects” of a face. Usually, such changeable aspects are contrasted with the invariant, static aspects such as the identity of a person. Given the discussion above of how motion can help recognition of identity in some cases, the idea of two *fully separate* streams seems unlikely in the light of these perceptual and developmental findings, however.

With the advent of fMRI as a widespread imaging technique, the question of how dynamic and static information about faces are dealt with in the brain has become the focus of a few recent studies. Again, however, the majority of studies have used non-natural, or highly abstract stimuli for testing any differences (cartoon faces, morphed expression-like stimuli, etc.). It seems clear from these earlier studies that, for example, the superior-temporal-sulcus (STS) in the dorsal pathway is active during perception of moving face stimuli (e.g., [2]). This region is also active during observation of biological motion and complex motion patterns such as optic flow in general [47].

Recently, two studies have directly contrasted activation in the brain during observation of static versus dynamic stimuli. In the first study [42], static and dynamic face stimuli were used in two different participant groups to localize areas involved in face perception. Such localizers are usually the first step in a fMRI study to identify candidate regions for closer inspection in the main part of the study. They found that using dynamic faces, face-sensitive regions could be much better identified than for static faces. Taking this one step further, a recent study [90] investigated response differences due to the dynamic information in the same group of participants. In addition to the expected activation in typical motion areas (Visual Area 5 and STS), they found that face regions in the ventral pathway that responded to static face stimuli were significantly more active for dynamic stimuli. This suggests that areas that are not traditionally associated with processing motion might already integrate dynamic information in the case of faces. At the very least, these results underline the fact that dynamic faces are the preferred stimulus for the brain.

22.2.3 *Summary*

In sum, it is clear that facial motion—both rigid and nonrigid—plays a role in the learning of new individuals and in the recognition of already learned individuals. This effect is strongest when the images are degraded. That is, the dynamic information helps to compensate for loss of static information. Since normal static images already contain a considerable amount of identity information, this is to be

expected for many reasons. It is not surprising that for a phenomena which has multiple sources of information, the removal of one source (e.g., static information) allows the effect of other sources (e.g., dynamic information) to be more clearly seen. Likewise, one might well imagine that the recognition of identity could be near ceiling performance for many types of task. Regardless, it is clear that not only do simple motions (such as horizontal rotations) help us to identify individuals, but that complex, characteristic motions (such as a certain way of smiling) provide distinct information about specific individuals. Future studies will need to clarify exactly what types of facial motions we remember about a person and how these might help us in identification. In addition, it seems that the beneficial effect of motion information is much more pronounced during early development of the perceptual apparatus, as the developmental studies have shown a clearer motion advantage also already for unfamiliar faces. These results highlight the fact that human face perception undergoes a long process of optimization and fine-tuning to let us become experts in face processing.

22.3 Dynamic Information for Expressions

Although less studied than identity perception, facial expressions are no less important for everyday life. Compared to other species, humans have developed highly sophisticated communication systems for social interaction. One of the most important of these is based on facial expressions. More specifically, facial expressions are known to serve a wide variety of functions:

- **Meaning:** They can, of course, be used to independently express complex ideas and intentions. For example, someone can look happy, sad, confused, or disgusted (see Fig. 22.2) [8, 9, 16, 29, 34, 62].
- **Modifier:** They are also very useful in *modifying* the meaning of auditory communication [11, 19, 25, 31, 76]. For example, a spoken statement by itself conveys a different meaning than when it is accompanied by a look of boredom. Indeed, in situations where the meaning conveyed by the face differs from that in another communication channel, the face tends to be considered more important [21, 41, 73].
- **Emphasis:** They co-occur with vocal accentuation [32, 79]. This emphasis can be seen, to some degree, in the static snapshot shown in Fig. 22.3.
- **Control:** Listeners can provide a wealth of information to the speaker without ever saying a word (this is referred to as “back-channel” signals; [111]). For example, a properly timed nod of agreement can tell the speaker to continue speaking, while a look of confusion at the same junction of the conversation would indicate that the speaker should stop and try to explain the last point again [10, 12, 18, 22, 23, 56, 86, 102].

Starting at birth, humans are trained to process faces and facial expressions, resulting in a high degree of perceptual expertise for face perception and social communication. This highly trained degree of expertise makes facial expressions—both

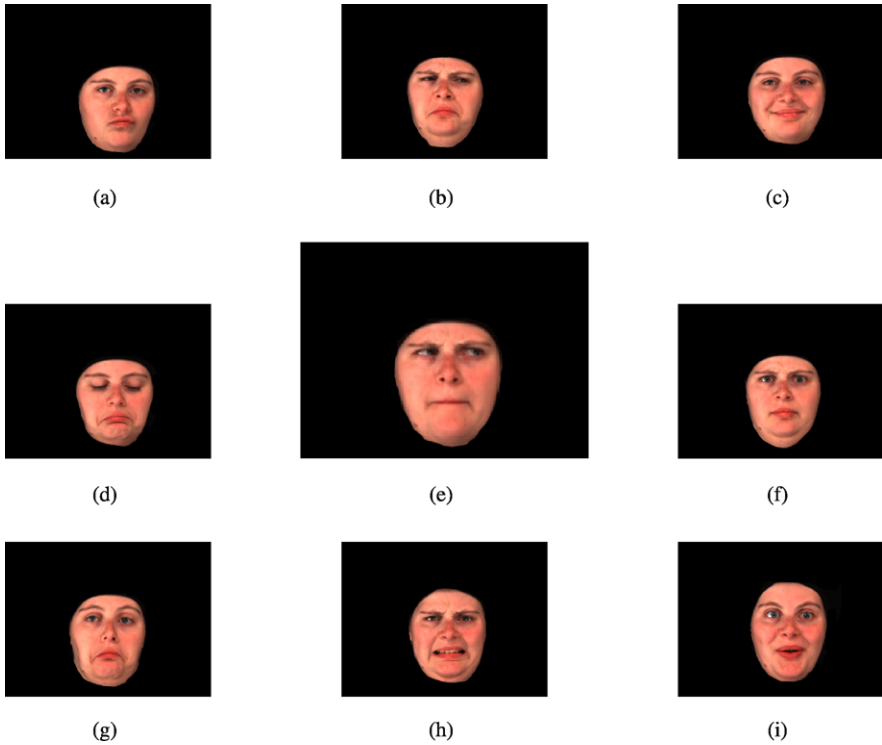


Fig. 22.2 Static snapshots of several facial expressions. **a** Agreement; **b** Disagreement; **c** Happiness; **d** Sadness; **e** Thinking; **f** Confusion; **g** Cluelessness; **h** Disgust; **i** Surprise. Some of these expressions can be recognized even in a static snap shot. Other expressions, like agreement and disagreement, seem to rely more heavily on dynamic information

emotional and conversational—an extremely difficult topic to study: We can detect very small differences in both motion and meaning. Furthermore, the physical differences between an expression that is recognizable (or is seen as sincere) and one that is not can be very subtle (see, for example, Fig. 22.4). Moreover, there are a number of different ways that humans express any given meaning, and not all of the resulting expressions are easily recognized [27–29]. All these factors jointly make the recognition of facial expressions one of the most difficult tasks the human visual system can perform [7].

Facial expressions have been the topic of scientific examination since at least [30]’s and [33]’s seminal work. These studies, as well as the majority of studies that followed, examined the reaction of people to various facial poses using photographs. Obviously, different facial areas are important for the recognition of different emotions: The mouth is critical for a smile, and the eyes for surprise, etc. [9, 29, 48, 78, 85]. For example, [37] showed that a true smile of enjoyment has not only the characteristic mouth shape, but also specific wrinkles near the eyes whereas faked expressions of enjoyment, in contrast, contain just the mouth information. This also

Fig. 22.3 A photograph of a facial expression that accompanied vocal emphasis. The facial expression that accompanies a vocal emphasis may be sufficient to recognize the emphatic nature of the statement, even in a static snapshot. This figure is taken from [29]



Fig. 22.4 Three photographs showing that small physical changes can produce large perceptual changes. These three snapshots were taken from three separate recordings of a clueless expression. All three expressions were recorded within 5 minutes of each other. This figure is adapted from [29]

shows that different facial regions can contribute to the perception of sincerity as well as to the recognition of the underlying expression.

The differential role of facial areas in different expressions is reflected in the fact that most models of facial expressions are explicitly parts-based [36, 38, 39, 43, 44, 57, 72, 101]. For example, Massaro and colleagues have 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 [38]. In one study, they used computer generated static facial expressions where either (a) the mouth, (b) the eyebrow, or (c) both were parametrically varied. Participants were asked to say if the expression was happiness or anger. Ellison and Massaro found that both features 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. There are, however, at least two models that integrate holistic (undecomposed whole face) information [58, 109].

Perhaps the most widely used method for *parameterizing* the high-dimensional space of facial expressions is the facial action coding system (or FACS, [36]), 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 is a system for describing the elements of photographs (or series of photographs) of facial expressions. It is not a model of facial expression processing per se and makes no claims in and of itself about which elements go together to produce different expressions [89].

Regardless of being parts-based, holistic, or hybrid, nearly all models of expression perception focus exclusively on static information. One can, sometimes include some information about the change of the static features over time by looking at the features in every frame, but there is rarely any ability to describe information that is only available over time. One potential exception to this rule is FACS+ from [40]. They used an optic flow technique combined with a few domain-based constraints to estimate facial structure and motion, yielding an empirical model of facial motion and structure. It is unclear, however, how the components of this model relate to the features used by humans. It is nonetheless increasingly clear that there is dynamic information for expressions and that any model of expression processing must take it into account.

Some of the earliest hints at spatiotemporal expression information comes from [8, 9], who used Johansson point-light faces to examine the role of motion in expression recognition (for more on point-light stimuli, see [59]). Their displays consisted of low-light recordings of the face and neck of several actors and actresses performing either an expression (happy, sad, surprise, disgust, interest, fear, and anger) or a random facial motion. The visible parts were painted black and then covered with approximately 100 white spots and the eyes were closed during the recording sessions. Thus, in the final video sequences, all that were visible were the 100 white points, moving about the screen. Each participant saw a single display (to avoid any learning effect or comparison of displays) and was asked to describe what they saw (such a free description task also helped to prevent biasing the participants’ answers). The display was either single static snapshot or a full video recording of one expression. On average, the collection of points was recognized as being a face considerably more often in the dynamic conditions than in the static conditions (73% versus 22% of the time, respectively). That is, the procedure removed nearly all static information that the display was a face as well as any information for more specific facial properties (such as identity or the specific expression). In a second experiment, an additional set of recordings where the face was visible (that is, no makeup was used) was included and participants were asked to identify the expression using a forced choice task (note that in a forced-choice task, a limited set of response options is given and the participant must choose one of these options as the answer). Overall, the expressions were recognized more often in the nonmakeup condition than in the point-light condition (65% versus 33% correct responses, respectively). Critically, more expressions were recognized in the point-light condition than can be expected by pure guessing, suggesting that there is some temporal information for facial expressions. Additional work suggests that even in dynamic

expressions, different expressions rely on different facial regions, with most expressions relying primarily on a single region [9, 29, 80]. Note that although eye, eyebrow, mouth, and rigid head motion are jointly sufficient for accurate recognition of the conversational expressions, other regions do contain information about facial expressions [80].

During the 20 years following Bassili's work, little to no studies examined the perception of dynamic facial expressions. In one of the few exceptions, [37] demonstrated that deceptive expressions of enjoyment appear to have different temporal characteristics than spontaneous ones. At the start of a rebirth in interest in dynamic expressions, [34] conducted an innovative experiment to demonstrate human sensitivity to temporal information. They printed out each frame of a video sequence of an expression. These photographs were given to participants (in a scrambled order), who were asked to place the photographs in the correct order. Participants were remarkably accurate, with 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.

Similar to [52]'s and [67]'s work with identity recognition, [62] examined the role of speed in the perception of expressions. Specifically, they manipulated the speed with which a neutral face transitioned into an emotional one. They found that happiness and surprise were better recognized from fast sequences, sadness better from slow sequences, and that angryness was best recognized at medium speeds. Subsequently, [53], demonstrated that increasing the distance traveled by an area while holding the timing constant (which also alters the speed and acceleration of the part) can exaggerate the emotional content of sentence. This suggests that different expressions seem to have a characteristic speed or rate of change.

Consistent with the work on identity, most examinations of dynamic expressions used degraded stimuli [3, 26, 35, 49, 60, 77, 105, 107, 108]. It has been consistently shown that dynamic expressions are recognized better than static ones. It has even been shown that the recognition of dynamic expressions with degraded static information can be as good as if not better than static expressions that are not degraded. That is, dynamic information can compensate for the loss of static information [35, 60, 105]. For example, [105] systematically degraded the shape, texture, and motion of a series of computer animated facial expressions. They examined performance in a forced-choice task, and found that dynamic sequences produce higher recognition rates than the static sequences. Likewise, they found that degrading either shape or texture information in the static conditions decreased performance. Critically, all dynamic conditions showed equal performance. That is, the presence of dynamic information eliminated the negative effect of degrading static information (specifically, shape and texture). In a separate experiment, they showed that animations that had proper nonrigid motion but lacked rigid head motion were recognized much worse than expressions that had both rigid and nonrigid motion. Additionally, the absence of rigid head motion greatly increased reaction times (participants were much slower in performing with rigid head motion). Finally, a simple,

temporal linear morph resulted in a small but significant drop in performance from the full motion condition, indicating that not only is motion important, but that natural motion seems to be important.

Just because one can use dynamic information does not, however, mean that one normally uses it. Indeed, in all the studies that have shown a dynamic advantage for expressions, the dynamic and static stimuli generally differed along a number of dimensions (such as the number of images and facial poses). Thus, it is unclear whether the dynamic advantage is due to (spatio)temporal information or something simpler. To help determine whether the dynamic advantage is due to the simple presence of more images, [3] compared recognition performance for a static expression, a dynamic version of that expression, and a condition where a 200 ms Gaussian noise mask was interspersed between the frames of the dynamic sequence. The noise was intended to mask the motion information. The normal dynamic condition resulted in much better performance than either of the other two conditions. Performance in the static condition and the masked dynamic condition did not differ from each other. This latter result confirms that masking does eliminate the perception of motion, as expected. Unfortunately, such a mask is also known to inhibit the processing of static information (see, e.g., backwards masking: [5, 61, 97, 110] or change blindness: [13, 87, 93]). Moreover, the stimuli contained only the early phases of an expression (i.e., the first three to six frames). This means that the static condition was degraded in a particular fashion: it did not contain all of the static information that is present in the peak expression which is used in other expression studies. Thus, it is not clear that the dynamic advantage found in [3]'s experiment would generalize to real-world situations.

Recently, [26] presented a series of 5 experiments conclusively demonstrating the presence of spatiotemporal information for facial expressions. Most of the experiments are strict analogues to the series run by [67, 69, 83] for facial identity. In Experiment 1, [26] directly compared dynamic and static peak versions of nine conversational expressions (see Fig. 22.2), seven of which demonstrated a dynamic advantage (happy and thinking were roughly equivalent in the static and dynamic conditions). This shows that a dynamic advantage can be found for video recordings over peak static images using normal intensity, conversational and emotional expressions of real individuals. The second experiment examined several static explanations for the dynamic advantages. For example, it is possible that the frame chosen in the first experiment was sub-optimal. Perhaps another frame was better, and the presence of this single image in the dynamic condition is the cause of the improved results. Likewise, since the perception of faces is, at least partially, based on its component parts, it is possible that people pick and choose the best parts from different frames, and composite them into a joint static whole (some evidence for this comes from [4]'s work with identity perception). Thus, similar to [69], a shortened dynamic sequence (the last 16 frames) and two static arrays (scrambled and ordered) were compared to the full dynamic and static peak conditions. The full dynamic and 16 frame dynamic conditions, which did not differ from one another, both produced higher recognition rates than the three static conditions (interestingly, the two array conditions produced slightly better performance than the static peak).

These results, combined with those from the third experiment (which scrambled the order of the frames in the dynamic sequence) show that the mere presence of many images or even face-appropriate dynamic information is *not sufficient*. There is some specific information present in the normal temporal development of an expression. Likewise, playing the expressions backwards (Experiment 4) reduced performance. Finally, the fifth experiment demonstrated that performance increases with increases in the number of frames that are kept together (blockwise scrambling). The length of the temporal integration window was at least 100 ms. In sum, dynamic expressions are recognized more easily and more accurately than static expressions, this effect is fairly robust, and the effect cannot be explained by simple, static-based explanations.

22.3.1 *Developmental Aspects*

Dynamic information is an even more crucial factor for processing of facial expressions during the perceptual development than it is for face identity [103]. Interestingly, despite a few early studies, there have been only relatively few recent studies that directly highlight the difference between static and dynamic processing of facial expression from a developmental perspective.

In an early study, [103] found that 5 and 7 month-old infants were able to discriminate between dynamically presented happy and angry facial expressions when they were presented with a congruent vocal expression. Following up on this finding, [95] used dynamic happy and angry expressions that were either point-light stimuli or normal faces. The study showed that infants, indeed, preferred the congruent stimuli over the incongruent ones in both conditions, highlighting the fact that the dynamic information in the visual and acoustic domains were integrated—this was especially true, of course, for the point light stimuli which presented much reduced shape information.

One of the most well-known perceptual findings about people suffering from autism spectrum disorder (ASD) is that they seem to have problems in identifying facial expressions. Interestingly, most research on this had been done with photographs, again presenting the participants with “unrealistic” stimuli. In a study comparing ASD and normal children on their performance with static and dynamic presentations of emotional facial expressions, surprisingly few differences between the two groups were found [45]. Interestingly, the authors hypothesized that it was the presence of slow dynamic changes in the stimuli that they had used that gave the autistic group a better chance at processing the stimuli—a result that was recently confirmed [98]. In another study, two groups of ASD and normal children were tested with a more complex set of facial expressions that were presented either statically or dynamically [6]. Although individuals with ASD performed significantly worse than the control group, there was little difference between static and dynamic presentation of expressions, which was most likely due to low, overall recognition performance (see [26]).

Finally, a recent larger, cross-sectional study tested an age range from 4 to 18 years with facial animations portraying different emotions at varying intensity [74]. The study found that the performance of expression recognition increased with age and that it also increased with the intensity with which emotions were animated. Interestingly, performance increased faster for the girls than the boys in the study indicating a gender difference during development of emotional understanding.

In summary, whereas the importance of dynamic information—and with that also the use of dynamic, natural stimuli—for the investigation of facial expression processing has been recognized, more studies are needed to elucidate the development of dynamic expression processing.

22.3.2 Neurophysiological Aspects

There is some emerging evidence that the neural mechanisms responsible for the perception of expression are at least partially different for static and dynamic facial expressions [1, 54, 66]. These studies include reports from patients who are completely unable to recognize expressions from static pictures or static descriptions, but have normal recognition performance for both dynamic expressions and descriptions of dynamic expressions or actions. A recent neural model for the processing of facial expressions was presented in [1], which details the different areas that might be involved at different points in time. Whereas this model accounts for some dynamic aspects, an integrated model of how the human brain interprets the highly complex dynamic signals from facial expressions is still lacking.

In one of the first studies to use real-world video sequences (rather than, for example, expression morphs such as in [66]), participants observed happy and angry facial expressions in both static and dynamic versions [63]. The study used Positron Emission Tomography (PET) to chart the different neural networks that were involved in perception of the different stimuli. In good agreement with the fMRI studies mentioned earlier, they found a series of typical motion areas to be activated for the dynamic stimuli. Additionally, they found critical differences depending on the expression used. That is, different networks of areas were associated with perception of dynamic happy expressions than with the perception of dynamic angry expressions, and those networks were in turn different from the networks found during perception of static expressions. A recent study [100] extended the stimulus material to include many more actors (thereby avoiding potential habituation effects as in previous studies, which mostly used expressions of only one actor/actress), and contrasted static and dynamic version of neutral, happy, and disgusted expressions using fMRI. Again, the results for the dynamic stimuli indicated a much more widespread network of activation than for the static stimuli. Interestingly, this also included pre-motor areas that are thought to be the human equivalent of the mirror-neuron system and perhaps might be related to motor imagery, or (unconscious) imitation of the observed expression. In addition, the recognition results for dynamic stimuli were found to be better than those for static stimuli [3, 26].

Whereas all studies seem to converge on the fact that facial expression perception in the brain for dynamic stimuli is different from that of static stimuli, a point of criticism still is that stimuli (and also the most prevalent existing models of expression processing [1]) in all cases are based on a few examples from the universal, emotional expressions. It remains to be seen how conversational facial expressions and thus more general facial movements will fit into the overall picture.

22.4 Conclusions

It is clear that there is some form of characteristic facial information that is only available over time, and that it plays an important role in the recognition of identity, expression, speech, and gender. It is also clear that the addition of dynamic information improves the recognizability of expressions and identity, and can compensate for the loss of static information. Moreover, at least several different types of motion seem to exist, which play different roles, and a simple rigid/nonrigid dichotomy is neither sufficient nor appropriate to describe. Additional research is necessary to determine what the dynamic features for face processing are.

The sole reliance on static information in any attempt to understand how humans use the face and head to identify or communicate with other people will illuminate only a very limited and maybe even artificial part of the perceptual and cognitive mechanisms involved. Likewise, any system designed to describe or recognize people or facial expressions that does not explicitly allow for the description of dynamic information will never yield human-like performance.

Artificial systems that aim at communicating naturally and efficiently with humans need to be based on a truly spatiotemporal description of the face and communication. Such a description will not only open the door for computer vision systems in biometrics and human-computer interaction, but also enable a more targeted analysis and a potential road to successful therapy and training approaches for patients with both deficits in communication production (such as occurring after a stroke, for example) or in communication understanding (patients with Autism Spectrum Disorder, or Asperger's syndrome, for example).

Acknowledgements We gratefully acknowledge the support of the Max Planck Society and the WCU (World Class University) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-2008-000-10008-0). We are grateful to useful discussions with those who read earlier drafts of the manuscript.

References

1. Adolphs, R., Tranel, D., Damasio, A.R.: Dissociable neural systems for recognizing emotions. *Brain Cogn.* **52**, 61–69 (2003)
2. Allison, T., Puce, A., McCarthy, G., et al.: Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* **4**(7), 267–278 (2000)

3. Ambadar, Z., Schooler, J.W., Cohn, J.: Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychol. Sci.* **16**, 403–410 (2005)
4. Anaki, D., Boyd, J., Moscovitch, M.: Temporal integration in face perception: Evidence of configural processing of temporally separated face parts. *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 1–19 (2007)
5. Averbach, E., Coriell, A.S.: Short-term memory in vision. *Bell Syst. Tech. J.* **40**, 309–328 (1961)
6. Back, E., Ropar, D., Mitchell, P.: Do the eyes have it? Inferring mental states from animated faces in autism. *Child Dev.* **78**(2), 397–411 (2007)
7. Barton, J.J.S.: Disorders of face perception and recognition. *Neurol. Clin.* **21**, 521–548 (2003)
8. Bassili, J.: Facial motion in the perception of faces and of emotional expression. *J. Exp. Psychol.* **4**, 373–379 (1978)
9. Bassili, J.: 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 (1979)
10. Bavelas, J.B., Black, A., Lemery, C.R., Mullett, J.: I show how you feel—motor mimicry as a communicative act. *J. Pers. Soc. Psychol.* **59**, 322–329 (1986)
11. Bavelas, J.B., Chovil, N.: Visible acts of meaning—an integrated message model of language in face-to-face dialogue. *J. Lang. Soc. Psychol.* **19**, 163–194 (2000)
12. Bavelas, J.B., Coates, L., Johnson, T.: Listeners as co-narrators. *J. Pers. Soc. Psychol.* **79**, 941–952 (2000)
13. Becker, M., Pashler, H.: Volatile visual representations: Failing to detect changes in recently processed information. *Psychon. Bull. Rev.* **9**, 744–750 (2002)
14. Berry, D.: What can a moving face tell us? *J. Pers. Soc. Psychol.* **58**, 1004–1014 (1990)
15. Berry, D.: Child and adult sensitivity to gender information in patterns of facial motion. *Ecol. Psychol.* **3**, 348–366 (1991)
16. Bruce, V.: *Recognising Faces*. Lawrence Erlbaum, Hillsdale (1988)
17. Bruce, V., Young, A.: Understanding face recognition. *Br. J. Psychol.* **77**, 305–327 (1986)
18. Bull, P.: State of the art: Nonverbal communication. *The Psychologist* **14**, 644–647 (2001)
19. Bull, R.E., Connelly, G.: Body movement and emphasis in speech. *J. Nonverbal Behav.* **9**, 169–187 (1986)
20. Calder, A.J., Young, A.W.: Understanding the recognition of facial identity and facial expression. *Nat. Rev., Neurosci.* **6**, 641–651 (2005)
21. Carrera-Levillain, P., Fernandez-Dols, J.: Neutral faces in context: Their emotional meaning and their function. *J. Nonverbal Behav.* **18**, 281–299 (1994)
22. Cassell, J., Thorisson, K.R.: The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Appl. Artif. Intell.* **13**, 519–538 (1999)
23. Cassell, J., Bickmore, T., Cambell, L., Vilhjalmsson, H., Yan, H.: More than just a pretty face: conversational protocols and the affordances of embodiment. *Knowl.-Based Syst.* **14**, 22–64 (2001)
24. Christie, F., Bruce, V.: The role of dynamic information in the recognition of unfamiliar faces. *Mem. Cogn.* **26**, 780–790 (1998)
25. Condon, W.S., Ogston, W.D.: Sound film analysis of normal and pathological behaviour patterns. *J. Nerv. Ment. Dis.* **143**, 338–347 (1966)
26. Cunningham, D.W., Wallraven, C.: Temporal information for the recognition of conversational expressions. *J. Vis.* **9**, 1–17 (2009)
27. Cunningham, D.W., Breidt, M., Kleiner, M., Wallraven, C., Bühlhoff, H.: The inaccuracy and insincerity of real faces. In: *Proceedings of Visualization, Imaging, and Image Processing* (2003)
28. Cunningham, D.W., Breidt, M., Kleiner, M., Wallraven, C., Bühlhoff, H.H.: How believable are real faces?: Towards a perceptual basis for conversational animation. In: *Computer Animation and Social Agents 2003*, pp. 23–29 (2003)
29. Cunningham, D.W., Kleiner, M., Wallraven, C., Bühlhoff, H.H.: Manipulating video sequences to determine the components of conversational facial expressions. *ACM Trans. Appl. Percept.* **2**(3), 251–269 (2005)

30. Darwin, C.: *The Expression of the Emotions in Man and Animals*. John Murray, London (1872)
31. DeCarlo, D., Revilla, C., Stone, M.: Making discourse visible: Coding and animating conversational facial displays. In: *Proceedings of the Computer Animation 2002*, pp. 11–16 (2002)
32. Dohen, M., Loevenbruck, Cathiard, M.-A., Schwartz, J.-L.: Audiovisual perception of contrastive focus in French. In: *Proceedings of the AVSP'03 Conference*, pp. 245–250 (2003)
33. Duchenne, B.: *The Mechanism of Human Facial Expression or an Electro-Physiological Analysis of the Expression of the Emotions*. Cambridge University Press, New York (1862/1990)
34. Edwards, K.: The face of time: Temporal cues in facial expressions of emotion. *Psychol. Sci.* **9**, 270–276 (1998)
35. Ehrlich, S.M., Schiano, D.J., Scheridan, K.: Communicating facial affect: It's not the realism, it's the motion. In: *Proceedings of the ACM CHI 2000 Conference on Human Factors in Computing Systems*, pp. 252–253. ACM Press, New York (2000)
36. Ekman, P., Friesen, W.: *Facial Action Coding System*. Consulting Psychologists Press, Palo Alto (1978)
37. Ekman, P., Friesen, W.V.: Felt, false, and miserable smiles. *J. Nonverbal Behav.* **6**, 238–252 (1982)
38. Elison, J.W., Massaro, D.W.: Featural evaluation, integration, and judgment of facial affect. *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 213–226 (1997)
39. Essa, I., Pentland, A.: A vision system for observing and extracting facial action parameters. In: *CVPR94*, pp. 76–83 (1994)
40. Essa, I., Pentland, A.: Coding, analysis, interpretation, and recognition of facial expressions (1997)
41. Fernandez-Dols, J., Wallbott, H., Sanchez, F.: Emotion category accessibility and the decoding of emotion from facial expression and context. *J. Nonverbal Behav.* **15**, 107–124 (1991)
42. Fox, C., Iaria, G., Barton, J.: Defining the face processing network: Optimization of the functional localizer in fMRI. *Hum. Brain Mapp.* **30**(5) (2009)
43. Frijda, N.H., Philipszoon, E.: Dimensions of recognition of emotion. *J. Abnorm. Soc. Psychol.* **66**, 45–51 (1963)
44. Frois-Wittmann, J.: The judgment of facial expression. *J. Exp. Psychol.* **13**, 113–151 (1930)
45. Gepner, B., Deruelle, C., Grynfeldt, S.: Motion and emotion: A novel approach to the study of face processing by young autistic children. *J. Autism Dev. Disord.* **31**(1), 37–45 (2001)
46. Gibson, J.J.: *The Ecological Approach to Visual Perception*. Lawrence Erlbaum, Hillsdale (1979)
47. Giese, M., Poggio, T.: Neural mechanisms for the recognition of biological movements. *Nat. Rev., Neurosci.* **4**(3), 179–192 (2003)
48. Hanawalt, N.: 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 (1944)
49. Harwood, N., Hall, L., Shinkfield, A.: Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. *Am. J. Ment. Retard.* **104**, 270–278 (1999)
50. Haxby, J.V., Gobbini, M.I., Furey, M., Ishai, A., Schouten, J., Pietrini, P.: Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001)
51. Hill, H., Johnston, A.: Categorization and identity from the biological motion of faces. *Curr. Biol.* **11**, 880–885 (2001)
52. Hill, H., Johnston, A.: Categorizing sex and identity from the biological motion of faces. *Curr. Biol.* **11**, 880–885 (2001)
53. Hill, H., Troje, N.F., Johnston, A.: Range- and domain-specific exaggeration of facial speech. *J. Vis.* **5**, 793–807 (2005)
54. Humphreys, G., Donnelly, N., Riddoch, M.: Expression is computed separately from facial identity, and is computed separately for moving and static faces: Neuropsychological evidence. *Neuropsychologia* **31**, 173–181 (1993)

55. Hunnius, S., Geuze, R.: Gaze shifting in infancy: A longitudinal study using dynamic faces and abstract stimuli. *Infant Behav. Dev.* **27**(3), 397–416 (2004)
56. Isaacs, E., Tang, J.: What video can and can't do for collaboration: a case study. In: *ACM Multimedia '93*, pp. 496–503. ACM, New York (1993)
57. Izard, C.E.: The maximally discriminative facial movement coding system (max). Available from Instructional Resource Center, University of Delaware. Newark, DE (1979)
58. Izard, C.E., Dougherty, L.M., Hembree, E.A.: A system for identifying affect expressions by holistic judgments (1983)
59. Johansson, G.: Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* **14**, 201–211 (1973)
60. Kaetsyri, J., Klucharev, V., Frydrych, M., Sams, M.: Identification of synthetic and natural emotional facial expressions. In: *International Conference on Audio-Visual Speech Processing (AVSP 2003)*, pp. 239–243 (2003)
61. Kahneman, D.: Method, findings, and theory in studies of visual masking. *Psychol. Bull.* **70**, 404–425 (1968)
62. Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., Akamatsu, S.: Dynamic properties influence the perception of facial expressions. *Perception* **30**, 875–887 (2001)
63. Kilts, C., Egan, G., Gideon, D., Ely, T., Hoffman, J.: Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *NeuroImage* **18**(1), 156–168 (2003)
64. Knappmeyer, B., Thornton, I., Bühlhoff, H.: The use of facial motion and facial form during the processing of identity. *Vis. Res.* **43**(18), 1921–1936 (2003)
65. Knight, B., Johnson, A.: The role of movement in face recognition. *Vis. Cogn.* **4**, 265–273 (1997)
66. LaBar, K., Crupain, M.J., Voyvodic, J.T., McCarthy, G.: Dynamic perception of facial affect and identity in the human brain. *Cereb. Cortex* **13**, 1023–1033 (2003)
67. Lander, K., Bruce, V.: Recognizing famous faces: exploring the benefits of facial motion. *Ecol. Psychol.* **12**, 259–272 (2000)
68. Lander, K., Chuang, L.: Why are moving faces easier to recognize? *Vis. Cogn.* **12**, 429–442 (2005)
69. Lander, K., Christie, F., Bruce, V.: The role of movement in the recognition of famous faces. *Mem. Cogn.* **27**, 974–985 (1999)
70. Lander, K., Bruce, V., Hill, H.: Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Appl. Cogn. Psychol.* **15**, 101–116 (2001)
71. Lander, K., Chuang, L., Wickam, L.: Recognizing face identity from natural and morphed smiles. *Q. J. Exp. Psychol.* **59**, 801–808 (2006)
72. Leventhal, H., Sharp, E.: Facial expression as indicators of distress. In: Tomkins, S.S., Izard, C.E. (eds.) *Affect, Cognition and Personality: Empirical Studies*, pp. 296–318. Springer, New York (1965)
73. Mehrabian, A., Ferris, S.: Inference of attitudes from nonverbal communication in two channels. *J. Consult. Psychol.* **31**, 248–252 (1967)
74. Montiroso, R., Peverelli, M., Frigerio, E., Crespi, M., Borgatti, R.: The development of dynamic facial expression recognition at different intensities in 4- to 18-year-olds. *Soc. Dev.* **19**(1), 71–92 (2010)
75. Mori, M.: Bukimi no tani. *Energy* **7**, 33–35 (1970) [The uncanny valley, (K.F. Macdorman and T. Minato, transl.)]
76. Motley, M.T.: Facial affect and verbal context in conversation—facial expression as interjection. *Hum. Commun. Res.* **20**, 3–40 (1993)
77. Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T., Vatikiotis-Bateson, E.: Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychol. Sci.* **15**, 133–137 (2004)
78. Numenmaa, T.: The language of the face. In: *Jyvaskyla Studies in Education, Psychology, and Social Research*. Jyvaskyla, Finland (1964). Jyvaskylan, Yliopistoydystys
79. Nusseck, M., Cunningham, D.W., Ruitter, J.P.D., Wallraven, C.: Perception of emphasis intensity in audio-visual speech. *Speech Commun.* 11 under review

80. Nusseck, M., Cunningham, D.W., Wallraven, C., Bühlhoff, H.H.: The contribution of different facial regions to the recognition of conversational expressions. *J. Vis.* **8**(8:1), 1–23 (2008)
81. O’Toole, A., Roark, D.: Memory for moving faces: The interplay of two recognition systems. In: Giese, M., Curio, C., Bühlhoff, H. (eds.) *Dynamic Faces: Insights from Experiments and Computation*. MIT Press, Cambridge (2009)
82. Otsuka, Y., Konishi, Y., Kanazawa, S., Yamaguchi, M., Abdi, H., O’Toole, A.: Recognition of moving and static faces by young infants. *Child Dev.* **80**(4), 1259–1271 (2009)
83. Pike, G., Kemp, R., Towell, N., Phillips, K.: Recognizing moving faces: The relative contribution of motion and perspective view information. *Vis. Cogn.* **4**, 409–437 (1997)
84. Pilz, K.S., Thornton, I.M., Bühlhoff, H.H.: A search advantage for faces learned in motion. *Exp. Brain Res.* **171**(4), 436–447 (2006)
85. Plutchik, R.: *The Emotions: Facts, Theories, and a New Model*. Random House, New York (1962)
86. Poggi, I., Pelachaud, C.: Performative facial expressions in animated faces. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.) *Embodied Conversational Agents*, pp. 115–188. MIT Press, Cambridge (2000)
87. Rensink, R.A., O’Regan, J.K., Clark, J.J.: To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.* **8**, 368–373 (1997)
88. Roark, D.A., O’Toole, A.J., Abdi, H., Barrett, S.E.: Learning the moves: The effect of familiarity and facial motion on person recognition across large changes in viewing format. *Perception* **35**, 761–773 (2006)
89. Sayette, M.A., Cohn, J.F., Wertz, J.M., Perrott, M.A., Parrott, D.J.: A psychometric evaluation of the facial action coding system for assessing spontaneous expression. *J. Nonverbal Behav.* **25**, 167–186 (2001)
90. Schultz, J., Pilz, K.: Natural facial motion enhances cortical responses to faces. *Exp. Brain Res.* **194**(3), 465–475 (2009)
91. Schwaninger, A., Wallraven, C., Cunningham, D.W., Chiller-Glaus, S.D.: Processing of facial identity and expression: A psychophysical, physiological and computational perspective. *Progress in Brain Research*, 325–348 (2006)
92. Schwarzer, G., Leder, H.: *The Development of Face Processing*. Hogrefe & Huber, Cambridge (2003)
93. Simons, D.J., Levin, D.T.: Failure to detect changes to people during a real-world interaction. *Psychon. Bull. Rev.* **5**, 644–649 (1998)
94. Skelton, F., Hay, D.: Do children utilize motion when recognizing faces? *Vis. Cogn.* **16**(4), 419–429 (2008)
95. Soken, N., Pick, A.: Intermodal perception of happy and angry expressive behaviors by seven-month-old infants. *Child Dev.* **63**(4), 787–795 (1992)
96. Spencer, J., O’Brien, J., Johnston, A., Hill, H.: Infants’ discrimination of faces by using biological motion cues. *Perception* **35**(1), 79 (2006)
97. Stigler, R.: Chronophotische Studien ber den Umgebungscontrast (Effects of exposure duration and luminance on the contrast of the surround). *Pflügers Arch. Gesamte Physiol. Menschen Tiere* **134**, 365–435 (1910)
98. Tardif, C., Laine, F., Rodriguez, M., Gepner, B.: Slowing down presentation of facial movements and vocal sounds enhances facial expression recognition and induces facial–vocal imitation in children with autism. *J. Autism Dev. Disord.* **37**(8), 1469–1484 (2007)
99. Thornton, I., Kourtzi, Z.: A matching advantage for dynamic human faces. *Perception* **31**, 113–132 (2002)
100. Trautmann, S., Fehr, T., Herrmann, M.: Emotions in motion: Dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Res.* **1284**, 100–115 (2009)
101. Tronick, E., Als, H., Brazelton, T.B.: Monadic phases: A structural descriptive analysis of infant–mother face-to-face interaction. *Merrill-Palmer Q. Behav. Dev.* **26**, 3–24 (1980)
102. Vertegaal, R.: Conversational awareness in multiparty vmc. In: *Extended Abstracts of CHI’97*, pp. 496–503. ACM, Atlanta (1997)

103. Walker-Andrews, A.: Infants' perception of expressive behaviors: Differentiation of multi-modal information. *Psychol. Bull.* **121**, 437–456 (1997)
104. Wallis, G., Bülthoff, H.H.: Effects of temporal association on recognition memory. *Proc. Natl. Acad. Sci. USA* **98**, 4800–4804 (2001)
105. Wallraven, C., Bülthoff, H.H., Fischer, J., Cunningham, D.W., Bartz, D.: Evaluation of real-world and computer-generated stylized facial expressions. *ACM Trans. Appl. Percept.* **4**, 1–24 (2007)
106. Wallraven, C., Breidt, M., Cunningham, D.W., Bülthoff, H.H.: Evaluating the perceptual realism of animated facial expressions. *ACM Trans. Appl. Percept.* **4**, 1–20 (2008)
107. Wehrle, T., Kaiser, S., Schmidt, S., Schere, K.R.: Studying the dynamics of emotional expressions using synthesized facial muscle movements. *J. Pers. Soc. Psychol.* **78**, 105–119 (2000)
108. Weyers, P., Mühlberger, A., Hefele, C., Pauli, P.: Electromyographic responses to static and dynamic avatar emotional facial expressions. *Psychophysiology* **43**, 450–453 (2006)
109. White, M.: Parts and wholes in expression recognition. *Cogn. Emot.* **14**, 39–60 (2000)
110. Williams, M., Breitmeyer, B., Lovegrove, W., Gutierrez, L.: Metacontrast with masks varying in spatial frequency and wavelength. *Vis. Res.* **31**, 2017–2023 (1991)
111. Yngve, V.H.: On getting a word in edgewise. In: *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pp. 567–578. Chicago Linguistic Society, Chicago (1970)