The effect of dynamics on identifying basic emotions from synthetic and natural faces

Jari Kätsyri*, Mikko Sams

Laboratory of Computational Engineering, Helsinki University of Technology, P.O. Box 9203, FIN-02015 HUT, Finland

Received 29 May 2007; received in revised form 31 August 2007; accepted 1 October 2007
Communicated by G. Lindgaard
Available online 7 October 2007

Abstract

The identification of basic emotions (anger, disgust, fear, happiness, sadness and surprise) has been studied widely from pictures of facial expressions. Until recently, the role of dynamic information in identifying facial emotions has received little attention. There is evidence that dynamics improves the identification of basic emotions from synthetic (computer-animated) facial expressions [Wehrle, T., Kaiser, S., Schmidt, S., Scherer, K.R., 2000. Studying dynamic models of facial expression of emotion using synthetic animated faces. Journal of Personality and Social Psychology 78 (1), 105–119]; however, similar result has not been confirmed with natural human faces. We compared the identification of basic emotions from both natural and synthetic dynamic vs. static facial expressions in 54 subjects. We found no significant differences in the identification of static and dynamic expressions from natural faces. In contrast, some synthetic dynamic expressions were identified much more accurately than static ones. This effect was evident only with synthetic facial expressions whose static displays were non-distinctive. Our results show that dynamics does not improve the identification of already distinctive static facial displays. On the other hand, dynamics has an important role for identifying subtle emotional expressions, particularly from computer-animated synthetic characters.

Keywords: Facial animation; Basic emotions; Movement perception

1. Introduction

Faces provide crucial information in social communication. Static facial features are important for identifying identity, gender and age from faces. Transient changes on face, driven by complex facial musculature, convey both verbal (speech) and non-verbal information. Facial expressions regulate turn-taking between speakers, emphasize speech, convey culture-specific signals and, importantly, reflect feelings of the speakers (Pelachaud et al., 1991).

A long research tradition suggests that at least six emotions (anger, disgust, fear, happiness, sadness and surprise) are ‘basic’ because they are identified distinctively from their characteristic facial expressions in all human cultures (Ekman et al., 1982). Although this claim has been debated, sometimes heatedly (e.g., Ortony and Turner, 1990; Ekman, 1994; Izard, 1994; Russell, 1994, 1995), several studies have supported the consistent identification of basic facial expressions in different cultures. Studies with a forced-choice task requiring participants to match presented facial expressions with a given list of emotion labels have confirmed that basic expressions are most often matched with intended basic emotions (e.g., Ekman, 1973, 1984; Ekman et al., 1982). Similar results have been found when participants have been asked to rate the intensity of each basic emotion in stimuli (e.g., Ekman et al., 1987) or to produce emotional labels for them freely (e.g.,...
Studies of basic emotions have typically utilized pictures of posed facial expressions such as the Ekman–Friesen collection of facial affects (Ekman and Friesen, 1978). The use of posed instead of authentic emotional facial expressions has raised questions on the ecological validity of such studies (e.g., Russell, 1994). Trivially, authentic everyday expressions are more natural than posed expressions. However, the latter ones do exhibit some advantages over the former. When using posed stimuli, actors can be trained in posing certain theoretically derived facial configurations exactly, producing homogeneous and distinctive emotional displays. On the other hand, authentic emotional expressions are more heterogeneous and their emotional content is typically ambiguous (cf. Ekman, 1973). For example, in a recently published collection of authentic basic expressions (O’Toole et al., 2005), several instances of happy facial displays were evoked with few instances of anger and fear (O’Toole, personal communication). Perhaps a more important issue than the use of posed instead of authentic emotional displays is the predominant use of pictures instead of moving emotional stimuli. Because most studies have been conducted with static facial pictures, the role of dynamic information on perceiving facial emotions has received little attention.

Previous studies have shown that dynamics (head movement and facial expression transitions) may have an important role in recognizing identity and age from faces. Identity recognition is to some extent possible from dynamic point-light displays (moving points) extracted from original faces of actors (Bruce and Valentine, 1988). Both identity and sex can be recognized when original facial movements are replicated on a computer-animated head showing none of the original static features (Hill and Johnston, 2001). These studies indicate that movement alone conveys some information about person’s identity and sex. Direct comparisons between static and dynamic displays have shown that observing movement enhances identity identification from faces when their presentation has been degraded by inversion, pixelation, blurring, luminance value thresholding or color negation (Knight et al., 1997; Lander et al., 1999, 2001).

There is evidence of the importance of dynamics also in identifying emotions from facial expressions. Studies with dynamic point-light displays have indicated that facial emotions can be identified from pure movement information (Bassili, 1978; Bruce and Valentine, 1988). Some neurological patients impaired in identifying emotions from still images of facial expressions do nevertheless recognize them from video sequences (Adolps et al., 2003) and point-light displays (Humphreys et al., 1993). By using schematic animations as stimuli, Wehrle et al. (2000) found better identification of dynamic rather than static displays of emotions. However, it is not clear whether this result was specific to synthetic facial stimuli because their results were not compared to dynamic vs. static natural facial expressions. Importantly, the synthetic static stimuli were identified worse than their static natural counterparts, suggesting that their result should not be generalized to natural faces. Studies using natural facial expressions have provided inconsistent results. Harwood et al. (1999) reported better identification of dynamic over static facial displays of emotions; however, such effect was observed only for anger and sadness. Kamachi et al. (2001) used image morphing to generate a video of a face changing from neutral to emotional, and found no difference between such dynamic and original static displays. Ehrlich et al. (2000) have reported better identification of basic emotions from dynamic rather than static facial expressions. However, because their results were pooled over good-quality facial expressions and their degraded versions, it is possible that the better identification of dynamic expressions was specific to the degraded stimuli. Recently, Ambadar et al. (2005) demonstrated that dynamics improves the identification of subtle, low-intensity, facial emotion expressions. However, full intensity facial expressions were not used as control stimuli. As a conclusion, dynamics appears to improve the identification of facial emotions from synthetic faces whose static presentations are not identified optimally. It is not established whether this effect applies also to full-intensity and non-degraded natural facial expressions.

The aim of the present study was to compare the identification of basic expressions from static and dynamic natural and synthetic faces in the same experiment. Natural stimuli consisted of posed, clearly distinguishable facial expressions of basic emotions. Synthetic stimuli were created with a three-dimensional head animation model (Frydrych et al., 2003). The used model did not capture all realistic facial details (cf. Section 2.2.1). Our hypothesis was that dynamics has no effect on the identification of already well-recognizable natural facial emotions but that it does improve the identification of those synthetic facial animations that are identified poorly from static displays.

2. Methods

2.1. Participants

Participants were 54 university students (36 males, 18 females; 20–29 years old) from Helsinki University of Technology (TKK) who participated in the experiment as a part of their studies. All participants were native speakers of Finnish and had either normal or corrected vision. The level of subjects’ alexithymic personality trait (Taylor et al., 1991) was evaluated with a Toronto Alexithymia Scale.
Alexithymia is defined as involving difficulties in expressing and experiencing emotions and an external orientation on own subjective emotional experiences. Alexithymia is also known to be associated with a reduced ability in identifying facial emotions (Parker et al., 1993; Mann et al., 1994; Lane et al., 1996). The TAS-20 scores of our subjects did not differ statistically significantly from the reference values of Finnish population (Salminen et al., 1999).

2.2. Stimuli

Synthetic stimuli were created with a three-dimensional animated “talking head” (TH) (Frydrych et al., 2003) developed in our laboratory. Dynamic natural stimuli were collected from an existing Cohn–Kanade (CK) collection (Kanade et al., 2000) of posed basic expressions. Additional dynamic stimuli were recorded specifically for this study (TKK stimuli). Because the CK and TKK stimuli had not been evaluated by naïve observers before, stimuli from a widely used Ekman–Friesen (EF) collection (Ekman and Friesen, 1978) served as control stimuli.

In total, the stimuli were six full basic facial expression sets (neutral, anger, disgust, fear, happiness, sadness and surprise) of static and dynamic natural (CK and TKK) and synthetic (TH) faces, and two control sets (EF) of static faces. Dynamic sets consisted of short (mean ± S.D.: 0.8 ± 0.2 s) video sequences (25 frames per second) showing transitions from neutral to emotional faces. Static sets consisted of original picture material or, if the originals were video sequences, last images from the video sequences. Stimulus size was either 22 cm × 17 cm (CK and TKK sets) or 14 cm × 20 cm (EF and TH sets).

2.2.1. Synthetic stimuli

Two synthetic sets of static and dynamic facial expressions were created with the TH animation software (Fig. 1). Stereotypical facial expressions for six basic emotions (Table 1) were defined in Facial Action Coding System (FACS) taxonomy (Ekman et al., 2002a) on the basis of existing literature (Ekman and Friesen, 1975; Ekman et al., 2002b) and were implemented manually on the animation model. The intensity of displays, i.e. the extent of changes between emotional and neutral facial expressions, was restricted by the current technical implementation based on rotational deformators allowing only relatively limited facial dynamics (cf. Frydrych et al., 2003). Dynamic facial expressions were extrapolated as a continuum between neutral and emotional faces. A certified FACS coder ensured that dynamic facial changes contained all of the intended FACS components. Two stimulus sets were created by using plain facial texture and by superimposing frontal and side photographs of a real person on the general mesh of TH by using an algorithm developed by Kulliomäki and Lampinen (2002). Superimposing a more
realistic facial texture produced purely aesthetical changes without any changes in the underlying facial expression model.

2.2.2. Natural stimuli

Two static/dynamic sets were selected from the CK collection (items 11-001, 11-004, 11-005, 14-002, 65-002 and 65-004; 14-004, 65-003, 66-001, 66-003, 71-002 and 71-004). An original FACS coding of the CK material was used to select such stimuli that resembled the TH basic expression stereotypes (Table 1) as closely as possible. Since it was not possible to find a suitable full set of all basic emotions from any single actor on this basis, such facial expressions were selected from various actors whose FACS codes were similar to the intended configurations. Consequently, the two sets contained stimuli from a total of five different actors.

Two static/dynamic sets were recorded in TKK from two certified FACS coders (initials JK and VK) (Ekman et al., 2002a) trained in controlling facial muscles associated with FACS Action Units. The TH basic expression stereotypes (Table 1) were used as basis for posing the facial expression configurations.

2.2.3. Control stimuli

Two static sets were selected from the EF collection, both sets containing pictures from one actor (items 1-04, 1-05, 1-14, 1-23, 1-30, 2-11 and 2-18 from actor MO; 2-05, 2-12, 2-16, 3-01, 3-11, 3-16 and 5-06 from actor WF). These items were selected on the basis of good recognition accuracy (88–100%) in the original EF evaluation study (Ekman and Friesen, 1978).

2.3. Procedure

The subjects were distributed randomly into two groups (n = 27), which were shown either static or dynamic stimuli sets. A between-subject design was used to avoid learning effects. As an exception, the control stimuli sets (EF) served as control stimuli for both groups. The static group evaluated also neutral faces in addition to the emotional expressions. “Static group” saw eight static sets of seven facial expressions (in total 56 stimuli) whereas “dynamic group” saw six dynamic sets and two static sets of six facial expressions (48 stimuli). The latter group evaluated six instead of seven different facial expressions because no dynamic versions of neutral stimuli were available. The TAS-20 scores did not differ significantly between the two groups.

Stimuli were presented with Presentation software (Version 0.53, www.neurobs.com). The order of stimuli was randomized. The subjects had to evaluate how well each of the six emotions (anger, disgust, fear, happiness, sadness and surprise) described the presented facial expressions. Evaluations were made on a scale from 1 to 7 (1 “totally disagree”, 4 “uncertain”, 7 “totally agree”) and given on a keyboard. In addition, subjects were asked to rate the naturalness of the stimuli using a similar scale (seven evaluations for each stimulus). Requests to evaluate were presented below the stimulus, one at a time in a random order. Each stimulus was shown until all seven (six basic emotion and naturalness) evaluations were made. There was no time limit for responses. Before the actual experiment, the subjects participated in a training session consisting of two randomly selected stimuli.

2.4. Analysis

2.4.1. Dependent variables

2.4.1.1. Identification accuracy. To analyze the identification accuracies with one dependent variable, we converted the six original emotion ratings to single identification scores. A similar approach has been utilized in earlier rating studies (e.g., Kamachi et al., 2001; Parker et al., 2005). The scoring method was devised to measure how distinctly the presented basic expressions were evaluated as depicting their intended basic emotions. When the intended basic emotion was rated higher than all other emotions, its identification was considered optimal and a maximum score of 1 was given. Each confusion, i.e. an unintended basic emotion receiving at least as high rating as the intended emotion, reduced the score by a constant of 0.4. For example, when there were two or three confusions, scores 0.6 (1–10.4), 0.2 (1–20.4), 0 (1–2.50.4) and −1 (1–50.4) were given. Ratings given at random would have produced an average of 2.5 confusions per evaluation and, respectively, a mean identification score of 0 (1–2.50.4). When the intended emotion was confused with all other emotions, i.e. it was rated no higher than any of the five remaining emotions, a minimum score of −1 was given (1–50.4).

2.4.1.2. Naturalness ratings. The original naturalness rating scale was re-expressed with standard transformation methods (e.g., Behrens, 1997) to meet normality assumptions required by parametric statistical tests. Because of the infrequent use of rating 4 (the “uncertain” answer), results for the original ratings 3 and 4 were pooled together. The obtained six-step scale was power transformed (with exponent 1.5) to reduce large negative skewness evident in the evaluation of several facial expressions. The re-expressed naturalness scale ranged from 1 to 14.7 (i.e., from 11.5 to 61.5).

2.4.1.3. Response latencies. For analyzing response latencies of evaluated stimuli, mean response latencies were calculated over their individual ratings.

2.4.2. Statistical analyses

Significance level was set to α = 0.01 for all analyses. All post-hoc comparisons were conducted with Newman–Keuls post-hoc tests.
3. Results

3.1. Effect of dynamics

A mixed-design ANOVA with factors Display (static, dynamic), Type (natural (CK and TKK), synthetic (TH)), and Expression (six basic expressions) was used with naturalness ratings and identification scores to evaluate the significance of the first factor and its interactions with the other factors. Control stimuli (EF) were excluded from this analysis as they were always static. With naturalness ratings, only the interaction between Display and Expression reached significance ($F(5,260) = 3.12, p = 0.01$). Analysis of simple effects indicated that dynamic happy expressions were evaluated less natural than static ones ($\text{mean} \pm \text{S.E.M.}: 6.6 \pm 0.4 \text{ vs. } 8.4 \pm 0.3; F(1,52) = 9.44, p = 0.004$). Further analysis with additional factor Actor (all CK, TKK and TH stimulus sets) indicated no significant differences between individual actors in the naturalness ratings of dynamic and static happy facial expressions ($F(5,260)<1$). With identification scores, the main effect of Display ($F(1,52) = 12.97, p < 0.001$) and the interaction Display $\times$ Type ($F(1,52) = 31.52, p < 0.001$) reached significance. Mean identification scores for static and dynamic synthetic and natural faces are depicted in Fig. 2. As clearly indicated by the figure, dynamics increased the identification of synthetic ($F(1,52) = 24.90, p < 0.001$) but not natural stimuli ($F(1,52)<1$).

Because response times were not controlled, the observed dynamics effect could have reflected a more careful evaluation strategy allowed by longer evaluation times. For evaluating this possibility, response latencies were analyzed with a mixed-design ANOVA with factors Display (static, dynamic) and Type (natural (CK, TKK), synthetic (TH)). A significant Display main effect confirmed that response latencies were significantly longer for dynamic rather than static stimuli ($\text{mean} \pm \text{S.E.M.}: 4.4 \pm 0.3 \text{ s vs. } 3.5 \pm 0.2 \text{ s}; F(1,52) = 8.39, p < 0.01$). However, response latencies were significantly longer for dynamic vs. static stimuli both with synthetic ($4.6 \pm 0.3 \text{ s vs. } 3.7 \pm 0.2 \text{ s}; F(1,52) = 8.90, p < 0.005$) and real faces ($4.2 \pm 0.3 \text{ s vs. } 3.3 \pm 0.2 \text{ s}; F(1,52) = 9.05, p < 0.005$), and a non-significant Display $\times$ Source interaction suggested that these differences were roughly equal. If the better identification of dynamic over static synthetic stimuli were due to longer response latencies, dynamic basic expressions should have been identified better also from real faces. Because this was not the case, it is unlikely that the dynamics effect of synthetic stimuli was related to response latencies. Plausibly, dynamic stimuli were evaluated longer than static ones because participants preferred to watch a full replication of dynamic video sequence before giving their answer.

The ANOVA for identification scores received significant Display $\times$ Expression ($F(5,260) = 9.65, p < 0.001$) and Display $\times$ Type $\times$ Expression ($F(5,260) = 10.16, p < 0.001$) interactions. Furthermore, the interaction between Display and Expression was significant with synthetic ($F(5,260) = 11.58, p < 0.001$) but not with natural faces ($F(5,260) = 1.21, p = \text{n.s.}$), indicating that dynamics effect differences between basic expressions occurred only with synthetic faces. Identification scores for static and dynamic displays of different synthetic expressions are shown in Fig. 3. Tests of simple effects showed significantly better identification of dynamic rather than static expressions of anger ($F(1,52) = 18.60, p < 0.001$) and disgust ($F(1,52) = 32.35, p < 0.001$).

3.2. Evaluation of synthetic faces

Synthetic facial expressions were identified worse than natural ones both from static ($F(1,52) = 153.65, p < 0.001$) and dynamic displays ($F(1,52) = 19.89, p < 0.001$) (cf. Fig. 2). The significance of identification score differences between synthetic and natural basic expressions were studied further with contrast tests,
conducted separately for static and dynamic displays (Table 2). The results showed that all static synthetic facial expressions except fear were identified significantly worse than their natural counterparts. With dynamic displays, fear and sadness were identified significantly and disgust close to significantly worse from synthetic rather than natural faces. The identification score of each static and dynamic synthetic expression was also compared with chance performance (zero score). With the exception of static anger (t(26)<1) and disgust (t(26)<0), all synthetic expressions were identified more accurately than chance (t(26)>2.8, p<0.005 (one-tailed)).

The significance of possible identification differences caused by the superimposed facial texture was studied with a mixed-design ANOVA with factors Texture (non/textured), Display (static, dynamic) and Expression (six basic expressions). The main effect of Texture and its interactions failed to reach significance (F<2.7, p = n.s.). Replicating the earlier results, significant main effects of Display (F(1,52) = 24.90, p<0.001) and Expression (F(5,260) = 15.31, p<0.001) and a significant interaction between them (F(5,260) = 11.58, p<0.001) were observed. Post-hoc analysis showed that, in general, angry (mean±S.E.M.: 0.39±0.08), disgusted (0.25±0.10) and fearful facial animations (0.39±0.07) were identified worse than the remaining animations, and sad animations (0.61±0.05) worse than surprised animations (0.87±0.02). The remaining differences were not significant.

3.3. Evaluation of natural faces

To study whether systematic differences between CK and TKK stimuli and EF control stimuli could have affected obtained results, a mixed-design ANOVA with factors Display (static, dynamic), Source (CK, TKK, EF) and Expression (six basic expressions) was conducted with identification scores. The main effect of Expression (F(5,265) = 38.69, p<0.001) and the interaction Source × Expression (F(10,530) = 5.12, p<0.001) were significant. Post-hoc analysis showed that in general, fearful faces (mean±S.E.M.: 0.60±0.03) were identified worse than all other expressions, angry faces (0.71±0.03) worse than the remaining expressions and sad faces (0.83±0.02) worse than happy (0.95±0.01) and surprised expressions (0.94±0.01). The remaining differences were non-significant.

Because the identification score differences between basic expressions were not significant between CK and TKK sources (F(5,260) = 2.01, p = n.s.), their results were pooled together and compared with those of EF stimuli. Analysis of simple effects showed that angry faces were identified significantly worse (0.65±0.04 vs. 0.83±0.03; F(1,52) = 9.62, p<0.004), and sad faces significantly better (0.89±0.03 vs. 0.72±0.05; F(1,52) = 11.70, p<0.002) from CK and TKK in comparison to EF faces. For evaluating the former result further, a new repeated-measures ANOVA with factor Actor (all six human actors) was conducted specifically with the identification scores of angry faces. The main result was significant (F(5,265) = 10.41, p<0.001), confirming that the identification of anger varied between actors. Post-hoc analysis was conducted in order to find those actors whose identification scores fell significantly below either one of the EF actors. The only such result was the worse identification of anger from TKK actor VK (0.35±0.10) in comparison to EF actors WF (0.79±0.06) and MO (0.87±0.05). When the result for this expression was removed, the difference in identification of anger was no longer significant between CK and TKK (0.75±0.03) in comparison to EF stimuli (F(1,52) = 1.83, p = n.s.). Removing this facial expression from other earlier analyses did not change their overall results.

4. Discussion

We studied the effect of dynamics on identifying six basic emotions from natural (human actors’) and synthetic (computer-animated) facial expressions. Our results showed no significant differences in the identification of static and dynamic expressions from natural faces. In contrast, dynamics increased the identification of synthetic
facial expressions, particularly those of anger and disgust. Although our static synthetic stimuli were identified generally worse than their natural counterparts, the static angry and disgusted expressions were the only expressions whose identification failed to exceed chance level. Apparently, the identification of basic emotions is already at ceiling from typical facial expressions posed by human actors (e.g., Ekman and Friesen, 1978), making dynamic information redundant. However, dynamics is of importance with non-distinctive static emotional facial displays, as was the case with our synthetic facial expressions of anger and disgust. These findings are congruent with previous studies showing that dynamics enhances the identification of facial identity only from degraded facial images (Knight and Johnston, 1997; Lander et al., 1999, 2001). Combining these results, it can be concluded that (i) dynamics does not enhance the processing of facial information under optimal viewing conditions, (ii) the role of dynamics becomes more important under compromised conditions, i.e. when the static features have been degraded and (iii) these conclusions apply both to the identification of identity and emotional expressions from faces.

The previous studies have shown that dynamics improves the identification of emotions from synthetic facial expressions (Wehrle et al., 2000) and from subtle (Ambadar et al., 2005) or degraded (Ehrlich et al., 2000) natural facial expressions. Ehrlich et al. (2000) studied both degraded and non-degraded natural stimuli; however, they did not report dynamics results separately for these two kinds of stimuli. Studies using non-degraded and full-intensity natural facial expressions as stimuli (Harwood et al., 1999; Kamachi et al., 2001) have provided inconsistent results. The present study builds upon these studies by comparing synthetic facial expressions and non-degraded natural facial expressions in the same experiment, and by showing that the effect of dynamics is restricted to the former.

The similar identification of static and dynamic natural facial expressions in our study is consistent with the earlier study by Kamachi et al. (2001) but inconsistent with the study by Harwood et al. (1999) where dynamics was found to improve at least the identification of anger and sadness. Because the static facial expressions used by Harwood et al. were identified as well as those selected from a previously evaluated collection (Mazurski and Bond, 1993), it is unlikely that the dynamics effect reported in their study was due to poor identification of static stimuli. The dynamic video sequences used in the present study were shorter (mean below 1 s) than those used by Harwood et al. (10 s). However, because video sequences were played in a loop until subjects gave their answers and were carefully evaluated (mean evaluation time was over 4 s for each basic emotion evaluation of a stimulus), it is highly unlikely that stimulus presentation time would have affected the results. On the other hand, the results could have been influenced by different kinds of movement information. Whereas in the present study video sequences showed a transition from a neutral face to an emotional apex with little or no rigid whole-head movements, the exact nature of movement in the study by Harwood et al. is uncertain. For example, extensive whole-head movements, obviously visible in dynamic but not static displays, could have enhanced the identification of dynamic emotional facial expressions disproportionally.

Similar identification results should be obtained for natural and synthetic stimuli if the latter ones are realistic enough. Comparison to standard facial expression stimuli (Ekman and Friesen, 1978) confirmed that the natural facial expressions used in the present study were excellent exemplars of the six basic emotions. In contrast, all static synthetic facial expressions except fear were identified significantly worse than natural ones. Identification of static anger and disgust, whose identification scores failed to exceed chance level, was especially poor. The poor identification of anger and disgust reflects the well-known fact that these two expressions are easily confused with each other even with natural faces (Ekman, 1973, 1984; Ekman et al., 1982, 1987; Rosenberg and Ekman, 1995; Haidt and Keltner, 1999). Even if dynamics facilitated the identification of synthetic facial expressions, at least fearful and sad synthetic facial expressions were identified worse than natural ones.

Two plausible explanations could be given for the worse identification of synthetic facial expressions. Firstly, as is the case with most animation models (e.g., Massaro, 1998; Wehrle et al., 2000); however, see also (Waters, 1987; Terzopoulos and Waters, 1993), the used animated face did not include realistic modeling of facial features such as skin wrinkling and bulging that are important for natural facial expressions (Ekman et al., 2002a). The lack of such static features could explain the poor identification of some of the used synthetic stimuli. This explanation is, however, made unlikely by evidence showing that the identification of facial expressions generated with non-detailed TH (Bartneck, 2001) or even with schematic line-drawings (Katsikitis, 1997) may reach or exceed that of real faces. More likely, the poor identification of the used synthetic facial expressions was due to their relatively low intensities (i.e., small changes between emotional and neutral faces) caused by technical limitations. Studies using both naturalistic (Hess et al., 1997) and schematic (Wehrle et al., 2000; Bartneck and Reichenbach, 2005) faces have shown that the identification accuracy of facial expressions decreases as their intensity is reduced. The fact that our natural and synthetic stimuli contained roughly similar FACS action unit changes confirms that their facial configurations resembled each other. However, the worse identification of static synthetic facial expressions suggests that their intensity levels were below those of natural stimuli. Apparently, synthetic angry and disgusted facial expressions in particular contained static cues that were too subtle to allow for correct identification. Interestingly, the effect of dynamics was pronounced with these two expressions. The fact that dynamics would improve the
identification of low-intensity facial expressions is consistent with the recent study by Ambadar et al. (2005) using extremely low-intensity facial expressions as stimuli.

In both our study and the study by Wehrle et al. (2000), synthetic facial expressions were identified worse than natural ones from static presentations. Both studies also found better identification of emotions from dynamic rather than static synthetic faces. Although the facial animation models used in our and Wehrle et al.’ studies have some obvious visual differences (e.g., the former is based on animating the full three-dimensional facial surface (Frydrych et al., 2003) whereas the latter is apparently based on animating only some facial contours (Wehrle et al., 2000)), they also share some important similarities. In both synthetic models, facial expressions were created manually on the basis of theoretically derived FACS configurations. The synthetic faces of Wehrle et al. appear to have been of rather low intensity in comparison to their natural counterparts (e.g., see Figs. 2 and 3 in Wehrle et al., 2000), as was the case in our study. Therefore, it is plausible that the better identification of dynamic rather than static dynamic displays in Wehrle et al.’ study was mostly due to the low intensity of static synthetic faces.

Although our main interest was in the identification accuracy of emotions, we also asked our participants to evaluate the naturalness of presented stimuli. A statistically significant dynamics effect was found in that dynamic happy facial expressions were considered less natural than their static versions. This effect was similar across the individual happiness expressions, including the synthetic ones. We are not aware of a similar result from previous work. In the present study, dynamic stimuli contained fast transitions from neutral to posed emotional configurations without any additional movements. Although such movement sequences are optimal in representing their intended emotions as distinctively as possible, they may also differ from spontaneous emotional expressions. It is possible that evaluators are hypersensitive to such differences with happy facial expressions because happy expressions are seen more often in everyday life than the other studied expressions. Arguably, the expression of happiness is suppressed less by cultural display rules than negative emotions (cf. Ekman, 1973).

The present study concentrated on studying the effect of dynamics on identifying emotional facial expressions. Obviously, motion perception is a general phenomenon restricted not only to faces or emotional facial expressions. For example, it is known that not only facial identity or facial emotions (Bassili, 1978; Bruce and Valentine, 1988) but also walking persons and even inanimate objects can be identified from pure movement information contained in moving point-lights (see Roark et al., 2003 for a review). Plausibly, the enhanced identification of dynamic emotional facial expressions observed in this study is at least partly due to the general capability of primate visual system for processing rapid changes in the environment in addition to its constant features. However, other factors specific to faces or emotional facial expressions could also be involved. Different hypotheses for the facilitating effect of motion in recognizing facial emotions have been studied recently by Ambadar et al. (2005). They managed to exclude explanations related to the larger amount of information contained in video sequences vs. single pictures and to the facilitation of configural processing, i.e. enhanced processing of relations between individual facial features (cf. Maurer et al., 2002). They also considered the existence of certain intermediate emotion-characteristic movements in dynamic video sequences. This explanation was discarded because presenting only the first and last frames from full video sequences produced as large dynamics effect as the original video sequences. The authors concluded that the facilitating effect of dynamics was due to enhanced change perception between emotional and neutral faces. This is a viable hypothesis; however, as their study used facial expressions that were too brief to contain intermediate emotion-specific movements (4–7 frames selected from the beginning of full-intensity emotional facial expressions), their conclusion is not fully supported.

The present study confirmed that dynamics does not improve the identification of basic emotions from already distinctive static facial displays, such as typical emotional expressions posed by human actors. In contrast, the results from a synthetic face suggested that dynamics does improve the identification of facial expressions involving subtle changes on the face. The results add to earlier knowledge on the perception of static facial expressions of basic emotions. On the other hand, the results should be interesting also in relation to human–computer interaction. In real communication between humans, subtle expressions probably have a more important role than high-intensity emotional facial expressions. For example, autistic individuals having severe social interaction deficits appear to identify basic expressions without difficulties but to have severe difficulties in understanding the emotional meaning of more subtle changes on the face (Baron-Cohen et al., 1997, 2001). Our results highlight the importance of subtle facial changes also on interaction between humans and computer-animated characters. A study by Bartneck and Reichenbach (2005) has shown that humans are able to recognize relatively subtle emotional facial expressions from the faces of animated characters. The evaluation of dynamic instead of static faces is closer to real interaction between agents (be they human or computer-generated) whose faces are constantly changing. The drastic difference between the identification of dynamic and static displays of some synthetic facial expressions shows that in actual interaction, some subtle facial expressions can be understood even if they were not recognizable from static pictures.

Acknowledgments

We thank Robotics Institute of Carnegie Mellon University for access to the Cohn–Kanade facial expression
database. We thank Michael Frydrych, Martin Dobšík and Vasily Klucharev for insightful discussions and for their concrete contributions to the present study. This study was supported partly by the Academy of Finland (Grant 213938 to M.S.; Finnish Centre of Excellence Programs 2000–2005 and 2006–2011, Grant nos. 202871 and 213470).

References


