Helsinki University of Technology Laboratory of Computational Engineering

Teknillisen korkeakoulun Laskennallisen tekniikan laboratorion julkaisuja

Espoo 2006 Report B59

HUMAN RECOGNITION OF BASIC EMOTIONS FROM POSED AND ANIMATED DYNAMIC FACIAL EXPRESSIONS

Jari Kätsyri

Dissertation for the degree of Doctor of Philosophy to be presented for public examination and debate in Auditorium K at Helsinki University of Technology on the 15th of December, 2006, at 12 o'clock noon.

Helsinki University of Technology

Department of Electrical and Communications Engineering

Laboratory of Computational Engineering

Teknillinen korkeakoulu

Sähkö- ja tietoliikennetekniikan osasto

Laskennallisen tekniikan laboratorio

Distribution:
Helsinki University of Technology
Laboratory of Computational Engineering
P.O.Box 9203
FIN-02015 TKK
FINLAND

Tel. +358 9 451 5735
Fax +358 9 451 4830
http://www.lce.hut.fi

Online in PDF format: http://lib.hut.fi/Diss/2006/isbn951228538X

E-mail: Jari.Katsyri@tkk.fi

ISBN-13 978-951-22-8537-2 (printed)
ISBN-10 951-22-8537-1 (printed)
ISBN-13 978-951-22-8538-9 (PDF)
ISBN-10 951-22-8538-X (PDF)
ISSN 1455-0474

Picaset Oy
Helsinki 2006

Abstract

Facial expressions are crucial for social communication, especially because they make it possible to express and perceive unspoken emotional and mental states. For example, neurodevelopmental disorders with social communication deficits, such as Asperger Syndrome (AS), often involve difficulties in interpreting emotional states from the facial expressions of others.

Rather little is known of the role of dynamics in recognizing emotions from faces. Better recognition of dynamic rather than static facial expressions of six basic emotions has been reported with animated faces; however, this result hasn't been confirmed reliably with real human faces. This thesis evaluates the role of dynamics in recognizing basic expressions from animated and human faces. With human faces, the further interaction between dynamics and the effect of removing fine details by low-pass filtering (blurring) is studied in adult individuals with and without AS. The results confirmed that dynamics facilitates the recognition of emotional facial expressions. This effect, however, was apparent only with the facial animation stimuli lacking detailed static facial features and other emotional cues and with blurred human faces. Some dynamic emotional animations were recognized drastically better than static ones. With basic expressions posed by human actors, the advantage of dynamic *vs.* static displays increased as a function of the blur level. Participants with and without AS performed similarly in recognizing basic emotions from original non-filtered and from dynamic *vs.* static facial expressions, suggesting that AS involves intact recognition of simple emotional states and movement from faces. Participants with AS were affected more by the removal of fine details than participants without AS. This result supports a "weak central coherence" account suggesting that AS and other autistic spectrum disorders are characterized by general perceptual difficulties in processing global *vs.* local level features.

Keywords: Social cognition, basic emotions, facial expressions, movement perception, facial animation, Asperger syndrome.

Tiivistelmä

Kasvonilmeet ovat tärkeä osa sosiaalista vuorovaikutusta, erityisesti koska ne tekevät ääneen lausumattomien tunnetilojen ilmaisemisen ja havaitsemisen mahdolliseksi. Esimerkiksi sosiaalisen vuorovaikutuksen ongelmia sisältäviin neurokehityksellisiin oireyhtymiin, kuten Aspergerin Syndroomaan (AS), liittyykin usein vaikeuksia kasvoilla näkyvien tunnetilojen tulkitsemisessa.

Liikkeen roolista tunneilmausten tunnistamisessa kasvoilta on olemassa vain vähän tietoa. On osoitettu, että dynaamiset perustunneilmaukset tunnistetaan staattisia paremmin tietokoneanimoiduilta kasvoilta, vastaavaa tulosta ei ole kuitenkaan varmennettu ihmiskasvoilla. Tässä väitöskirjassa tutkitaan liikkeen roolia perustunneilmausten tunnistamisessa animoiduilta- ja ihmiskasvoilta. Ihmiskasvojen tapauksessa tutkitaan vuorovaikutusta liikkeen ja alipäästösuodatuksen (sumennuksen) kautta tapahtuvan tarkkojen yksityiskohtien poistamisen välillä. Tätä kysymystä tutkitaan lisäksi erikseen henkilöillä, joilla ei ole viitteitä AS:sta ja henkilöillä joilla on todettu AS. Tulokset vahvistivat, että liike edesauttaa tunneilmausten tunnistamista kasvoilta. Tämä tulos oli kuitenkin havaittavissa vain käytetyillä kasvoanimaatioilla, joista puuttui kasvojen tarkkoja yksityiskohtia ja muita tunteisiin liittyviä vihjeitä sekä sumennetuilla ihmiskasvoilla. Jotkin dynaamiset tunneanimaatiot tunnistettiin huomattavasti staattisia paremmin. Ihmisnäyttelijöiden esittämien perustunneilmausten tapauksessa, liikkeen tuoma lisähyöty kasvoi käytetyn sumennustason funktiona. Osallistujat, joilla oli todettu AS, tunnistivat perustunneilmauksia yhtä hyvin alkuperäisiltä ei-sumennetuilta kasvoilta ja dynaamisilta vs. staattisilta kasvoilta kuin muutkin osallistujat. Tulokset antavat viitteitä vahingoittumasta yksinkertaisten tunneilmausten ja liikkeen tunnistamisesta kasvoilta Aspergerin Syndroomassa. Osallistujat, joilla oli AS, suoriutuivat muita osallistujia heikommin, kun esitetyistä ärsykkeistä oli poistettu tarkkoja yksityiskohtia. Tämä tulos on yhdenmukainen ”heikoksi keskeiseksi koherenssiksi” nimetyn näkemyksen kanssa, jonka mukaan AS:aan ja muihin autismin kirjon häiriöihin liittyy havaitsemistason vaikeuksia yleisten vs. tarkkojen piirteiden prosessoinnissa.

Asiasanat: Sosiaalinen kognitio, perustunteet, kasvonilmeet, liikkeen havaitseminen, kasvoanimaatio, Aspergerin Syndrooma.

Author: Jari Kätsyri, M. Sc.
Laboratory of Computational Engineering
Helsinki University of Technology
Finland

Supervisor: Academy Professor Mikko Sams
Laboratory of Computational Engineering
Helsinki University of Technology
Finland

Preliminary examiners: Veikko Surakka, Assistant Professor, Ph. D.
Department of Computer Sciences
University of Tampere
Finland

Risto Näsänen, Docent, Ph. D.
Brainwork Laboratory
Finnish Institute of Occupational Health

Official opponent: Professor Jari Hietanen
Department of Psychology
University of Tampere
Finland

List of abbreviations

AS	Asperger syndrome
ASD	Autistic spectrum disorder(s)
AU	Action unit
Basic expression	Facial expression of basic emotion
FACS	Facial Action Coding System
FACSAID	FACS Affective Interpretation Dictionary
SDD	Social developmental disorder

Preface

This thesis concludes my work in the Laboratory of Computational Engineering at the Helsinki University of Technology during years 2001-2006. My work has been funded by Helsinki Graduate School in Computer Science and Engineering (HeCSE) and by a grant from the Emil Aaltonen's Foundation.

I wish to express my deepest gratitude to academy professor Mikko Sams for the expert guidance I've received along the way. It is no large exaggeration to state that without professor Sams, the present thesis wouldn't have become reality.

I am genuinely grateful to my official reviewers Dr. Risto Näsänen and Dr. Veikko Surakka who gave me both encouraging feedback and well put constructive criticism.

I thank the people who collaborated in the studies of this thesis, both for providing invaluable support and enlightening discussions: Prof. Lennart von Wendt (study V), Drs. Martin Dobšík (study II), Michael Frydrych (studies II-IV), Vasily Klucharev (study III) and Kaisa Tiippana (studies IV-V) and M.Sc. Satu Saalasti (study V). Basic implementation of "TKK talking head" facial animation engine was provided by Dr. Frydrych, Dr. Dobšík, M.Sc. Andrej Krylov and M.Sc. Pertti Palo. Dr. Frydrych also provided digital image filters required in studies IV and V. I'm thankful to the actors who participated in our facial expression recordings. I wish to express my thanks to the students and volunteers who literally made our experiments possible.

Laboratory of Computational Engineering, and especially its cognitive science and technology research group, has provided an inspiring multidisciplinary atmosphere for research work. I wish to thank my colleagues Laura Kauhanen, Jussi Kumpula, Ville Lilja, Riikka Möttönen, Tapio Nieminen, Ville Ojanen, Johanna Pekkola, Cajus Pomrén, Toni Tamminen, Iina Tarnanen, Riitta Toivonen, Mikko Viinikainen, and many others for our co-operation and interesting discussions.

I wish to thank my sister Seija and my good friends Toni, Saku and Mikko for their friendship and intellectual stimulation extending beyond academic circles. Most importantly, I wish to thank Kati who has shared both my good and bad times during the making of this thesis.

Jari Kätsyri

Contents

1 INTRODUCTION

Faces are crucial for social communication. Faces carry information about the identity, sex and age of their owners. Facial movements have several roles in conversation. Visible articulatory movements are known to enhance and influence the perception of speech. Facial actions punctuate and emphasize speech (*e.g.* brow movements), convey signals of their own that typically depend on culture (head nodding and shaking and eye winking), and regulate turns during speech (changes in head position and eye gaze) [1, 2]. Importantly, facial expressions also allow an access into the internal emotional and mental states of others.

According to a classical emotion theory, there are six or seven *basic emotions*, shared by all people regardless of their origin [3]. Such a conclusion has been supported by studies showing that people across the world from Westerners to members of isolated tribes are able to recognize these emotions readily from stereotypical facial displays (*e.g.* [4]). These and later studies on the recognition of emotions from human faces have until recently been conducted almost exclusively with photographs of facial expressions with only a few studies using moving faces as research stimuli. Consequently, there exists relatively little information on the role of motion in recognizing emotions from faces. The lack of emotion studies with moving facial expressions is partly explained by the lack of available stimuli. A picture collection of basic expressions by Ekman and Friesen [5] was collected already in the 1970's and has since been used widely in emotion research. Video sequence collections comparable to this collection have been non-existent until recently and their availability remains scarce.

The role of motion in recognizing basic emotions from facial expressions forms the main research question of this thesis. To evaluate whether earlier results with facial animations [6] can be generalized to real human faces, the question is studied both with emotional facial animations and emotional facial expressions posed by human actors. The main hypothesis is that dynamics improves the recognition of basic expressions, but only when static stimuli are degraded. This hypothesis is studied directly with the posed emotional facial expressions that have been filtered to produce various levels of blurred research stimuli. The recognition of basic emotions from static and moving, degraded and

non-degraded facial expression stimuli is compared further between neurotypical persons and persons with Asperger syndrome (*e.g.* [7]), a neurocognitive disorder belonging to an autistic spectrum disorders [8] and characterized by deficits in using and understanding non-verbal communication. The remainder of this introduction considers the roots of basic emotion theory and the existing knowledge on other topics considered in this thesis. An overview of conducted studies is given at the end of this discussion.

1.1 Theory of basic emotions

Basic postulations

The historical roots of basic emotion theory originate from Charles Darwin, who as a part of his evolutionary theory suggested that the emotional expressions of man were descendants from other animals [9] (cf. [4: pp. 169-75]). Darwin not only made observations on the behavior of animals but also set to study the question of whether some emotions were universal to all men. Although the idea of universal basic emotions had been mentioned already many centuries before Darwin in the writings of philosophers such as Descartes, Hobbes and Spinoza [10] and influential facial expression studies had been conducted by other 19th century scientists such as Guillaume Duchenne [11], Darwin appears to have conducted the first scientific evaluation studies on the recognition of emotions from faces. Darwin studied which emotions were recognized consistently from photographs of representative emotional facial expressions in England [9] (cf. [4: pp. 169-75, 12]) and made the first attempts to evaluate the universality of emotions by interviewing his fellow countrymen living abroad on the expression of emotions in other cultures [4: pp. 169-75]. The method of asking subjects to judge emotions from certain facial expressions has remained a part of contemporary research methodology.

After Darwin, several researchers have set to study basic emotions. According to a critical review [13], different basic emotion sets ranging from two to eighteen basic emotions have been proposed by different investigators; however, most of them agreeing at least on emotions *anger, fear, happiness* and *sadness*. One of the most influential researchers on emotional facial expressions and the theory of basic emotions has been Paul Ekman, who has suggested that the six emotions *anger, disgust, fear, happiness,*

sadness and *surprise*, and possibly *contempt*, are basic [3, 4, 14, 15]. Ekman's postulations on emotions are discussed here because they have had a significant influence on contemporary views and because they have been, at least to large extent, supported by empirical evidence. The evidence for the suggested six basic emotions is based on emotional facial expression judgment studies conducted in Western and isolated cultures. In a study by Ekman and Friesen [4: pp. 204-8], subjects from five different countries selected the expected emotion most often from a written list of six basic emotions when asked to evaluate photographs of facial expressions supposedly characterizing the six basic emotions. The original result was confirmed by other similar studies in total of 21 countries [16]. It is important to note that although the subjects in these studies were provided with a predefined list of the alleged basic emotions, this list wasn't selected arbitrarily (cf. critique by Russell [17: pp. 116-8]). In a later review by Ekman, Friesen and Ellsworth [3], it was stated that the alleged six basic emotions had been found in nearly all of the earlier studies. In some of these studies, the six basic emotions had been selected consistently from a very large number of response options (up to 100 emotion categories).

It was possible that all of the foregoing studies by Ekman and coworkers were influenced by common visual experience provided by Western mass media, such as television, movies and magazines. To eliminate this possibility, a new study was conducted in New Guinea within an isolated tribe called the Fores [4: pp. 204-8, 18]. This study was repeated independently by Heider and Rosch (described in [4: p. 214, 19: p. 713]) in a similar New Guinean tribe the Danes. Because most of the participants were illiterate, they were asked to select which one of the three presented pictures of emotional facial expressions best depicted an emotional story. Importantly, results from Fores indicated no differences between subjects whose contact with Western culture was extensive (who had lived in a Western settlement, received at least one year of education, seen movies and so on) and those whose contact was nonexistent or minimal, suggesting that having contacts with Western culture doesn't affect the results of an emotion judgment task. In general, all basic expressions were recognized as expected by at least another one of the groups; however, the Fores often chose surprised face for a fearful story and the Danes disgusted face for an angry story [19: p. 713]. Ekman has suggested

that these misattributions were due to display rules [18: p. 71] (see more below); however, the results do raise doubts on the universal recognition of surprised and disgusted emotions from their characteristic facial expressions. When emotions posed by Fores were shown to American subjects, they recognized basic emotions from them as predicted, except for fear and surprise whose recognition was intermixed. The problems with fear and surprise resemble misattributions made by the Fores in recognizing emotions; however, also the poses may have been rather poor as the Fores were inexperienced actors and puzzled by the task [4: p. 212]. An additional problem in the studies was that the used three response options didn't always include all relevant emotional facial expressions. For example, the Fores weren't given a possibility for selecting an angry face in response to a disgust-related story [18] (cf. [17]) which could have been a potential further misattribution.

Contempt has been suggested later as a basic emotion in addition to the six original ones and it has received some support from a study conducted in ten countries [15]; however, no studies have been made in isolated groups. Furthermore, subjects have failed to recognize contempt from supposedly characteristic facial expressions in more recent studies [20, 21], indicating that contempt shouldn't be considered a basic emotion.

The studies in Western cultures appear to provide strong evidence and the studies in isolated cultures at least partial evidence for the six basic emotions. Granting that this evidence is sufficient, it is interesting to consider the relation between basic and other emotions, as the spectrum of possible affective states is clearly much wider than six distinct emotions. Ekman has suggested that more complex emotions can be created by blending basic emotions, for example smugness could be thought of as a blend of happiness and contempt (assuming contempt as a basic emotion) [22, 23]. The concept of blended emotions is compelling; however, no extensive rules have ever been defined for forming non-basic out of basic emotions (cf. [13: p. 326]) (simple pair-wise facial expression blends have been described in [24]). It has also been suggested that basic emotions should be thought of as *emotion families* covering various related emotions rather than individual emotions [25]. This suggestion drastically increases the number of emotions covered by basic emotion theory. The concept of emotion families is given some support for example by a study which used hierarchical analysis to study subjects'

similarity ratings between common emotion words [26]. As a result, the emotion words were categorized under broad clusters resembling the six basic emotions proposed by Ekman, except for *disgust* that didn't exist as a cluster and *love* that did although not being one of the proposed basic emotions. The evidence from this study is limited because it was conducted only in one culture.

Conceptual and methodological problems

The basic emotion theory proposed by Ekman has been criticized because of various conceptual and methodological problems (see especially [13, 17]).

Attribution vs. expression of emotions

Reasoning behind the emotion judgment studies used to evaluate basic emotion theory is that certain emotions are basic because they are recognized universally from certain facial expressions. However, universal *attribution* of certain emotion labels to certain facial expressions doesn't necessarily imply the universal *expression* of the referred emotions by those facial expressions (cf. [13, 27]). In order to prove the latter, it would be necessary to show that certain facial expressions accompany certain emotional events universally. On the contrary, it appears that similar events may lead to different emotions and emotional expressions in different cultures [4: pp. 174-87]. How could such variation be compatible with the existence of basic emotions? According to a *neuro-cultural theory of emotions* suggested by Ekman [4: pp. 175-9, 22: pp. 212-35], certain basic emotions are universal but their eliciting situations and *display rules* controlling their expression differ between cultures. Display rules have been claimed to work by intensifying, deintensifying, neutralizing or masking the emotional facial expressions [22: pp. 212-35]. Before studying the cross-cultural expression of emotion in a certain situation, one would need to show that the situation elicits a similar emotion in the studied cultures and that no display rules affecting the results would be evident.

A study by Ekman and Friesen with spontaneous facial expressions (reported in [4: pp. 214-20]) has supported the existence of display rules. In their study, the faces of Japanese and American subjects were videotaped while watching stressful and neutral films. Earlier evidence had suggested that the Japanese and Americans produce similar emotional self-reports when watching the selected films [ibid]. The influence of display

rules was made less probable by videotaping the subjects without their explicit knowledge and by controlling whether the subjects were alone or accompanied by another person. When another person was present in the room, the Japanese showed more positive and less negative emotions than the Americans. This supported the researchers' hypothesis that the Japanese would conceal their negative reactions while not being alone because of the effects of cultural display rules. Analysis of the videotapes also showed a strong correlation in the emotional facial expressions between the groups, giving some support for the universal expression of emotions by certain emotional facial expressions. This evidence was, however, limited to expressions of fear and disgust, as these appear to have been the only evoked emotions (cf. [19: p. 713]).

The relation between basic emotions and facial expressions

The emotion judgment studies also assume that certain facial expressions are characteristic of certain basic emotions. This assumption raises further questions on the relation between facial expressions and emotions. Silvan Tomkins has suggested the face as "[…] of the greatest importance in producing the feel of affect" [28: p. 212]. Tomkins had a strong influence on Ekman's thinking, who has for example expressed doubts on the very concept of facial *expression* of emotion: "[…] in my view expression is a central feature of emotion, not simply an outer manifestation of an internal phenomena" [25: p. 384].

Can facial expressions occur without emotions? Some facial expressions certainly do occur without any emotion because only a minority of all facial expressions are related to emotions (cf. [29: p. 337]). A less trivial question is whether *emotional* facial expressions can occur without emotion. Ekman suggests that some facial muscles related to genuinely felt emotions are extremely difficult to activate voluntarily [25: pp. 389-91], for example the activation of *orbicularis oculi* (a circular muscle surrounding the eye) may discriminate genuine from non-genuine smiles. Furthermore, some tentative evidence exists on the fact that consciously posing emotion-related facial expressions can produce slight emotional experiences [30] (as cited in [25: p. 50]). Of course, facial expressions alone should not be expected to induce strong emotions because emotions (and facial expressions) are usually triggered in real or imagined situations.

Can emotions occur without facial expressions? Ekman's original postulations suggested that cultural display rules may interrupt the *visible* activation of facial muscles [22: pp. 212-35]. Respectively, emotion-related facial expressions may sometimes be inhibited so that felt emotions aren't accompanied with any visible facial activation [14, 25]. However, Ekman has also suggested that when emotions occur, impulses are always sent to facial muscles even if the actual activation would be interrupted later by display rules or conscious control [29]. This suggestion has been supported by findings on that even when emotions aren't accompanied with visible facial expressions, electrodes attached on the face reveal sub-visible facial muscle activity [14, 25].

It appears that in Ekman's view certain emotional states are always associated with certain facial muscle activations (*i.e.* facial expressions) and vice versa, although also other factors are involved. Because of the effect of display rules, the facial muscle activations characteristic for different emotions are not always observed as such in the emotional situations of real life. Facial expressions used in emotion judgment studies are "prototypical" in a sense that they should represent as pure emotional expressions as possible without any influence from display rules or other factors. In fact, it is no exaggeration to claim that the existence of such prototypical basic expressions is crucial for the basic emotion theory. For being able to describe such prototypes, an objective language for facial actions is needed. The most recent facial expression classification tool is FACS (Facial Action Coding System) by Ekman, Friesen and Hager [31], which is intended for describing all visually detectable changes on the face produced by facial muscle activity. The FACS system is used by human observers, after extensive training, to recognize and classify subtle facial actions. Facial actions are described with objective and emotion-independent *action units* (AUs), depicted with abbreviations such as AU1, AU4 or AU1+4. An evaluation study has shown that action unit coding with FACS has good to excellent inter-observer reliability [32].

Considering the importance of prototypical expressions for the basic emotion theory, it is disappointing that no unequivocal FACS descriptions have ever been defined for them (the lack of such definitions has been noted also by Ekman, cf. [3: p. 39]). Furthermore, it appears that no FACS coding is readily available for a collection of pictures of facial affects by Ekman and Friesen [5] containing a large number of photographs used in their

cross-cultural studies. The authors of FACS have given several suggestions for action unit combinations prototypical to basic expressions [33: pp. 173-5]; however, these suggestions are tentative instead of definitive. The predecessor of the FACS system, FAST (Facial Affect Scoring Technique) [33: pp. 101-32], was designed to describe directly facial actions related to basic emotions and contained also definitions for prototypical facial expressions. These prototypes were made obsolete after FACS replaced the FAST system. Simplified instructions for coding only emotion-relevant facial actions have been published in EMFACS [33: pp. 136-7, 34], which however doesn't contain suggestions on basic emotion prototypes. A digital dictionary called FACSAID (FACS Affect Interpretation Dictionary) [35] contains emotional interpretations for various action unit combinations observed in research and clinical practice, but doesn't define prototypes for basic expressions.

Ecological relevance

According to different arguments related to ecological critique, it is questionable whether the results obtained with posed and rather exaggerated emotional facial expressions detached from any context can be generalized to emotional situations observed in everyday life. In everyday life, the emotional meaning of facial expressions often depends on context. The basic emotion theory has been criticized, for example, because polite smiles may occur without actual feeling of happiness and a facial expression of distress/sadness may occur with an intense positive emotion [13: p. 321]. These examples can be explained rather trivially by concepts already introduced earlier. The existence of polite smiles is well compatible with the effects of cultural display rules and polite smiles may be discriminated from genuine ones by the lack of specific muscle activations. In the case of intense positive emotions, simplistic assumptions are made on what situations elicit what emotions. It could be argued that the intensity of felt emotion itself may cause distress – sad or distressed facial expressions may after all reflect the felt emotion correctly even in a happy situation.

More serious problem is that contextual information may change the emotional interpretation of facial expressions. For example, Carroll and Russell [36] found that the recognized emotion depended on the background story presented together with identical facial expressions. This indicates that the emotional interpretation of facial expressions is

affected by other information, but doesn't necessarily contradict the existence of basic emotions. Ekman has suggested several types of information that can be recognized from facial expressions, including the preceding context, the expresser's cognitive processes, his physical state and his subsequent actions, and a word describing his emotion [29]. In a typical situation in everyday life, all these information types are either known or being evaluated. A difficulty related to studying the expression of emotions in real situations, as mentioned earlier, is that the initiation and expression of emotion may be affected by culture. Concerning the possible cultural variation, it may have been rather reasonable to constrain the emotion judgment studies on only one component (facial expressions) of emotion. As stated by Ekman, emotion judgment studies using context-detached facial expressions are used to study "the question of what the face *can* signal, not what information it typically *does* signal" [25]. Although not explicitly studied, culture and language can be expected to have an effect on the use of emotion words. Consequently, the consistency across cultures in attributing emotion words to facial expressions suggests that the face alone *can* signal information about basic emotions independently of the observer's culture. Even so, the finding that context affects the emotional interpretation of facial expressions makes it clear that basic emotion theory in its simple form is insufficient for explaining how emotions arise in natural situations. A more comprehensive theory on the initiation and processing of emotions is outside the scope of this thesis.

Stimuli in most cross-cultural studies have been posed facial expressions of emotion. It has been claimed that it is questionable whether such results can be generalized to judgment of emotions from spontaneous facial expressions in natural situations [17: pp. 114-5]. Ekman has suggested that if posed expressions used in the cross-cultural studies were totally different from spontaneous ones, the consistence of results across cultures would be difficult to explain [18]. For example, not only did New Guineans understand most of the emotions posed by Westerners but also Westerners understood most of the emotions posed by New Guineans. Of course, Ekman's reasoning begs the question by assuming that the posed expressions were evaluated similarly because they *do* represent genuine expressions of emotion. However, this assumption appears to be the best explanation available. As noted by Ekman [ibid], to explain the cross-cultural

agreement without assuming some similarity between posed and spontaneous expressions would require that all studied cultures would for some inexplicable reason have learnt to associate similar emotions with the evaluated posed expressions.

In considering the critique of using posed emotional facial expressions further, it is important to consider how the expressions were selected in the studies of Ekman and his coworkers. First of all, not all of the facial expressions used in the original cross-cultural study by Ekman and Friesen were posed [22], as the research material was selected from over 3000 photographs of both posed and spontaneous emotional expressions used in earlier studies [4]. Furthermore, such stimuli were selected that contained facial actions assumedly related to genuine emotions as defined by FAST basic expression prototypes (see above), regardless of whether they were originally posed or spontaneous. Consequently, it is too simplistic to state that the stimuli consisted of posed instead of authentic emotional facial expressions. On the contrary, prior hypotheses based on earlier research were made on which facial actions best depict genuinely felt emotions, and the stimuli were selected based on these hypotheses.

Gradient of recognition

Haidt and Keltner [21] studied recently the evaluation of various basic and other emotions from facial expressions in American and Indian subjects. Subjects were asked to evaluate both the emotional content in presented pictures and eliciting situations leading to those emotional expressions. Both types of evaluations were combined into composite scores measuring simultaneously i) how distinctively the emotional expressions were recognized and ii) to what extent the two studied cultures agreed in evaluating them. Studied emotions were found to form a linearly decreasing "gradient" between the most and the least recognized emotions. The prior hypothesis postulated on the basis of basic emotion theory was that a clear distinction would be observed between universally recognizable and non-recognizable emotions (*e.g.* between basic and other emotions). The fact that a linear decrease was observed instead of this clear-cut distinction, has been labeled as the "gradient critique".

The observed gradient for the recognition of different basic expressions between two cultures complements earlier findings on cultural variation in the recognition accuracies of basic expressions [17]. It has been claimed that such variation is acceptable as long as

the intended basic emotions are recognized more often from basic expressions than other emotions (see more below) in all studied cultures [37]. This suggestion appears reasonable and, if accepted, makes the gradient critique of Haidt and Keltner irrelevant. Furthermore, the composite score used in their study appears problematic because it combines agreement of the situation that elicited a facial expression with agreement of which emotion the facial expression depicted. This is problematic because, as suggested by the neuro-cultural theory presented earlier, certain emotional facial expressions may be universal although their eliciting situations and display rules controlling their expression may vary between cultures. For example, although Haidt and Keltner found that American subjects considered anger to be caused by violations of the self's rights more often than Indians, both Americans and Indians might still recognize anger similarly from typical angry faces.

The composite score also appears to confound cultural differences in the evaluation of basic expressions with non-cultural differences in their distinctiveness. A certain c*onfusion* pattern is known to exist between basic expressions [3, 4, 16, 19, 22: p. 266, 38]. Confusions refer to instances where a facial expression is judged consistently as some other emotion than the expected one[1]. Happiness is practically never confused with any other emotion whereas confusions are more evident between the negative emotions, especially between fear and surprise (fearful faces are often judged as surprised), and disgust and anger (disgusted faces judged as angry). It has been found that subjects are consistent in judging the second most common emotion from facial expressions, further suggesting that the confusion pattern is systematic [19, 22: p. 266]. The existence of systematic confusions instead of unambiguous recognition is slightly problematic for the basic emotion theory; however, it doesn't appear to threaten the validity of Ekman's studies, as long as the predicted emotions are recognized the most often and above chance level [37]. This obviously was the case in Ekman's original study in five literate cultures [4: pp. 204-8], because the percentage of predicted responses was over 70% for each basic emotion, ranging from 77% for fear to 95% for happiness (chance level for

[1] Because unexpected emotions are always recognized in addition to the expected emotion from certain basic expressions, *i.e.* related to the evaluated stimuli instead of observers making the evaluations, the term "confusion" is misleading (cf. [39]). However, the term is adopted here because of its wide use.

selecting one response out of six would have been approximately 17%). The confusion pattern probably reflects perceptual similarities between the basic emotions: happy facial expressions differ more from negative emotions than negative emotions do from each other, as already noted by Darwin [37: p. 271]. This suggestion has been supported by a computer vision study, where an algorithm made similar confusions in classifying emotional facial expressions from visual data as humans even if it had no prior knowledge about the emotions themselves [40].

Forced-choice response format

A common methodological critique (cf. [17: pp. 116-123]) of basic emotion studies is related to the use of forced-choice paradigm where subjects are asked to pick the most representative choice from a predefined list of emotion words. It is possible that different results would be obtained if the subjects were given a possibility for providing their own emotion labels freely.

Free-response studies have typically produced contradictory results (see [17, 37, 41] for reviews). A common observation has been that the expected emotions are selected somewhat less often in free-response than in forced-choice studies, possibly because a free-response task is both more demanding for the subject and more difficult to analyze [37, 41]. How accurately the predicted emotions are recognized and whether they are selected the most often seems to depend on how the results are categorized. For example, Russell [42] found that angry faces were labeled frustrated (31%) more often than angry (26%). When also anger-related synonyms were considered acceptable, a greater proportion of subjects selected anger (41%) than frustration, but more than a half of the subjects still selected an unpredicted emotion label. In his response to Russel's critique, Ekman claimed that frustration should be considered as a correct answer because it is generally considered as an antecedent for anger [37]. This being, the majority (71%) of subjects selected the predicted emotion. In a similar free-response study by Rosenberg and Ekman [43] where a systematic classification method suggested by the authors was used, all the basic emotions were recognized above chance-level and by the majority of subjects. Also the recent study by Haidt and Keltner [21] indicates that free-response method produces consistent recognition results for the six originally suggested basic emotions (and additionally for embarrassment).

Alternative theories

The fact that the intended basic emotions are recognized above chance from pictures depicting them gives support for the theory of basic emotions but, as noted by Russell [17, 44], it doesn't rule out other alternative hypotheses that would produce similar results in the same task, such as componential and dimensional theories of emotion. If observers evaluated the stimuli initially in terms of emotional components or emotional dimensions, the results would remain quantitatively the same.

According to componential theories, there is a universal relation between emotional components and components of facial expressions (see [13] for an example). For example, frown might be universally related to frustration and compression of the lips to determination. The componential approach may be congruent with the existence of basic emotions. At least frustration and determination could be grouped under the emotion family of anger. Prototypical facial expressions occur rarely as such in everyday life; their components might instead reflect emotional states related to basic emotions or their blends. Furthermore, the fact that Ekman and coworkers have suggested various facial expression prototypes for each basic emotion [33: pp. 173-5] instead of unequivocal prototypes suggests that they implicitly accept a componential view.

According to dimensional theories (see [3, 45, 46] for reviews), emotions can be reduced to a low-dimensional emotion space such as that spanned by *pleasantness* and *attention-rejection* as suggested by Woodworth and Schlosberg [47] (as cited in [45]) or *pleasantness* and *arousal* as suggested more recently by Russell [46]. First such theories were not based on quantitative analyses; however, later theories have utilized factor analysis to reduce original data to a small number of dimensions. The data has for example included sorting results of emotional terms, similarity ratings between pairs of facial expressions and results from emotion judgment studies. The dimensional theory offers some improvements to the basic emotion theory. Because all possible emotions are located in a common emotion space, the relation between non-basic and basic emotions is not problematic. The dimensions themselves may be based on confusion data from emotion judgment studies, thereby inherently explaining the existence of confusions between basic emotion categories.

In a dimensional model, each basic (or other) emotion would be located in the space spanned by the underlying emotional dimensions. For example, anger and happiness are both thought to represent neutral value on the dimension attention *vs.* rejection but opposite negative and positive pleasantness values in the two-dimensional model by Woodworth and Schlosberg [47] (as cited in [45]). Young and co-workers [45] used computer morphing for generating continuums of blended images between all pair-wise combinations of six representative basic expressions. If the evaluation of basic expressions were based on dimensions such as pleasantness and attention-rejection, a continuum between two basic expressions should cover all emotional states falling between their respective positions in the emotion space. For example, the continuum between angry and happy expressions should cover anger (negative valence), neutral emotional state (neutral valence) and happiness (positive valence). Such a pattern wasn't observed with any continuum between basic expressions. On the contrary, all of the blended images were perceived categorically, *i.e.* classified as either one of their componential basic emotions with an abrupt shift between them in the continuum. Respectively, the results by Young *et al.* give support for categorical instead of dimensional view on emotions.

Conclusion

The concept of universal basic emotions has deep historical roots and its basic postulations appear to have survived a critical scientific discussion. Some potential shortcomings of the basic emotion theory were introduced. Most notably, the original evidence from isolated aboriginal groups is inconclusive because of misattributions related to surprised and disgusted faces, and unequivocal facial expression prototypes for basic emotions haven't been defined despite of their importance for the basic emotion theory. On the other hand, the original emotion judgment studies in Western cultures and more recent studies with refined research methods have provided strong evidence for the existence of at least six universally recognizable basic emotions. Although other views on emotions, such as componential and dimensional theories, are compatible with these results, the basic emotion theory as suggested by Ekman appears to remain a viable starting point for emotion research.

1.2 Research methods used in emotional facial expression studies

Research material

High-quality stimulus material is necessary for emotional expression research. It would be laborious to prepare facial expression stimuli for each specific study. It is more common to utilize existing facial expression collections. The selection of research material is guided by theoretical and practical research needs. For example, whether one assumes a dimensional or categorical theory of emotions may affect what kind of emotional states the facial expressions should depict. In the present thesis, the main interest is in recognition of emotions from facial expressions of basic emotions containing no other information on the face (such as speech).

An important methodological decision is whether to use posed or spontaneous facial expressions of emotions (cf. Chapter 1.1). In both cases, the material should contain realistic, easily recognizable and unambiguous displays of basic expressions. It is probably easier to obtain realistic spontaneous than posed emotional facial expressions, but spontaneous facial expressions may also be more prone to unwanted artifacts. For example, display rules may cause some intended emotional facial expressions to be masked (*e.g.* when smile is used to mask an otherwise angry facial expression [24: pp. 107-12]), de-intensified or neutralized [22: pp. 212-35]. Creativity is required for inventing emotional stimuli for eliciting as pure instances of all basic emotions as possible. When recording spontaneous facial expressions, the camera should be concealed to reduce the effect of display rules (cf. [4]). This leads to possible ethical problems.

It has been suggested that posed expressions are approximate but possibly exaggerated forms of spontaneous facial expressions [4: pp. 180-5]. Posed facial expressions are probably less ambiguous and easier to recognize than spontaneous ones, but natural facial expressions of genuine emotions are difficult to pose. On the other hand, the recognizability and naturalness of posed facial expressions depend also on a further distinction between posed facial expressions. Ekman has suggested a distinction between emblematic and simulated posed expressions, where the former are used intentionally to

signal that the emotion is only referred to but not felt and the latter to convince that the emotion is felt (whether it actually is or not) [4: pp. 179-85]. This distinction is important because when non-professional actors are asked to pose emotions without further instructions, the facial expressions may be emblematic instead of simulated. The emblematic facial expressions should be avoided, because they differ from genuine emotions to emphasize the fact that an emotion isn't felt. It is suggested here that the typical simulated poses can be divided further into FACS-based and empathized expressions. The method of training actors to produce emotion-related FACS action unit configurations originates from Ekman and Friesen [5]. When successful, the recorded stimuli are uniform and resemble genuine emotional facial expressions. Training requires certified FACS specialists and is obviously both difficult and time-consuming. Instead of posing certain facial expressions, the actors could be asked to produce facial expressions by empathizing the asked emotions. At best this kind of facial expressions may be close to spontaneous ones, especially when recorded from professional actors trained in emotional self-elicitation methods such as Stanislawski technique [48] (as cited in [49]) – if empathizing emotions weren't possible, the acting profession would be difficult indeed.

Evaluation methods

After the research material has been collected, it is necessary to evaluate how closely the recorded facial expressions depict the intended emotions. If a large sample has been collected, evaluation can be used to select the most suitable facial expression samples. Two kinds of evaluation approaches can be used: emotion judgment studies by subjects and facial action coding by specialists. Assuming that the collected facial expression material has been confirmed to depict the intended emotions closely enough, emotion judgment studies by subjects are of course suitable for other emotion perception studies.

Facial action coding with FACS is especially useful if some predictions are made on which action units are related to expressions of genuine emotions. After the FACS coding with emotion-independent action units is completed, it is possible to evaluate whether all action units relevant for the intended emotions are present and whether the stimulus contains other confounding facial actions. Typically, three kinds of tasks have been used in emotion judgment studies: *forced-choice*, *rating* and *free-response* tasks [27] (cf.

Chapter 1.1). In forced-choice task, subjects are given a list of emotion words and asked to classify each facial expression as one of the given options. Rating task resembles forced-choice task, but in it subjects are asked to evaluate the intensity or existence of all of the given emotions, for example on a seven-step rating scale. Typically, six basic emotions are used as response options or evaluation dimensions in forced-choice and rating studies. In free-response tasks the subjects are asked to produce their own labels freely, often with the further constraint that the words should be emotion-related. Some minor variations of these methods are also used. For example, in a forced-choice study the subjects could be allowed to select several response options or asked both to select the best response and to rate its intensity. Ekman has suggested that in order to distinguish emblematic from simulated expressions, the subjects should be asked to judge not only what emotions are posed but also whether the actor is actually feeling the posed emotion or not [4: p. 191].

It is suggested here that rating method is in general more suitable than forced-choice method for the evaluation of basic expressions because of the existence of common confusions (cf. Chapter 1.1). Rating data on all basic emotion dimensions provides conclusive information about confusions made by each individual subject whereas forced-choice data provides such information only indirectly as proportions of subjects selecting unexpected emotions. Respectively, rating data requires a smaller subject sample and may reveal more subtle confusions than forced-choice data. Information about confusions is important because it can be used when evaluating whether changes in recognition accuracy were due to changed recognition of the expected emotion, unexpected emotions or both. Although confusions typically are a nuisance in the analysis of emotion judgment studies, they can also be studied explicitly. For example, cross-cultural consistency in the recognition of the most common unexpected emotion has been studied by Ekman and coworkers [19] in a rating study and more recently by Elfenbein and coworkers [50] in a forced-choice study.

It is also suggested that forced-choice and rating methods are as suitable for most purposes as the free-response method. Although not accepted by all researchers [17], earlier studies appear to confirm that subjects produce emotion labels similar or related to basic emotions in free-response studies [21, 41, 43] (cf. Chapter 1.1). Free-labeling

method could be expected to be appropriate at least for cross-cultural facial expression studies related to the elicitation of emotion (cf. [21]). In other studies, forced-choice and rating studies with predefined emotion labels can be expected to be both faster and less tiresome for subjects and easier to analyze than free-response studies.

Analysis

There are several methodological difficulties in analyzing data from emotion judgment studies. Difficulties related to forced-choice and rating tasks are discussed here.

The most straightforward way for analyzing forced-choice data is to use simple hit rates as a recognition score, *i.e.* to calculate the proportion of subjects judging a facial expression as the expected emotion. Wagner has suggested that such analysis suffers from response bias [27, 51], *i.e.* tendency of some subjects for recognizing certain emotions from any facial expression. For example, if in an extreme case a subject would classify all facial expressions as representing anger, he or she would receive inflated recognition scores when actually recognizing anger from an angry facial expression. Contrary to the assertion by Wagner [51: p. 10], the use of rating instead of forced-choice task doesn't appear to solve the problem because quantitative ratings would be as prone to emotional bias as categorical decisions. Wagner has suggested response bias to have little significance in the evaluation of easily recognizable emotional facial expressions [51: p. 9]. He has proposed the use of an *unbiased hit rate* (H_U) calculated by formula $H_U = H * (1-F)$ where H refers to hit rate and F to false alarm rate (the proportion of subjects judging other facial expressions than intended as depicting a certain emotion) (cf. [50: p. 7]). In practice, the hit rate for recognizing an emotion from a facial expression is corrected for how often the emotion label is used with other expressions. A potential problem here is that the unbiased hit rate doesn't make a distinction between subjective biases and stimulus-related confusions. For example, the fact that anger is always recognized to some extent from disgusted facial expressions (cf. Chapter 1.1) doesn't imply that subjects would be generally biased to recognizing anger. Reducing the recognition scores for angry expressions just because disgusted expressions look angry appears to be misguided. An analogical case would perhaps be considering red color as being less red because orange color resembles red color. It is suggested here that how

often certain emotions are recognized depends more on such confusions related to basic expressions than on subjective emotion biases. To author's knowledge, this hypothesis hasn't been tested.

The nature of subjective ratings is a potential problem for rating studies. The intervals between consecutive steps within a subjective rating scale, often called Likert scale, aren't necessarily equal. For example, on a 5-step Likert scale, the real difference between steps 1-2 (*e.g.* no emotion – slight emotion) isn't necessarily equal to that between steps 2-3 (*e.g.* slight – moderate emotion). Likert scale should respectively be considered as ordinal instead of interval scale on a traditional measurement scale hierarchy [52: pp. 17-21]. Consequently, statistical parameters such as mean and standard deviation, direct arithmetic operations and parametric statistical tests assuming interval scale variables shouldn't been used with Likert scale variables. However, this conclusion is a controversial one. It has been suggested, for example, that analyzing Likert scale variables with parametric statistical tests will produce *sufficiently* accurate results [53: pp. 61-3]. In fact, it has been common practice in emotion judgment studies to analyze ratings with parametric statistical tests (*e.g.* [54]), to normalize ratings across subjects (*e.g.* [55]) or to conduct arithmetic operations on ratings (*e.g.* [54, 56, 57]).

In rating studies, each facial expression is evaluated on several emotional dimensions. Because analyses based on all possible ratings would be difficult to interpret, data reduction is necessary. A straightforward method is to analyze only the ratings for the predicted target emotions (cf. [55]). In this method, a majority of ratings (those for other than the target emotions) are ignored. Instead of selecting only ratings for target emotions, all ratings for evaluated stimuli could be transformed into recognition scores. A simple approach is to consider ratings successful only when the predicted target emotion is rated higher than all other emotions and to assign scores 1 and 0 accordingly. Most of the available information is omitted also in this method. In some studies, situations where some emotions have been given as high ratings as the target emotion are taken into account in the scoring. In an evaluation study by Ekman, such *ties* were simply dropped from analysis, resulting in the omission of 5% of evaluations [5]. It appears that in such approach, the results are improved artificially. Alternatively, the recognition score could be decreased when ties are observed. For example, Kamachi *et al* [55] used a recognition

score of 1 divided by the number of ties (for example, if a disgusted expression was rated as high in anger and disgust, the score was 1/2 = 0.5). It appears that such scoring is far from an interval scale because the score is decreased the less the more ties there are (for example, intervals between consecutive scores 1/2, 1/3 and 1/4 would be approximately $1/3-1/2 \approx 0.17$ and $1/4-1/3 \approx 0.08$) Furthermore, the scoring leads to some unintuitive cases. For example, if all emotions were given the same rating, a score of 0.2 would be obtained (1/5=0.2); whereas if the target emotion were rated one point lower than one other emotion but higher than all others, the score would be zero. In a study by Parker and coworkers [56], basic expressions were evaluated on 5-step Likert scales related to all basic emotions and the recognition score was formed by dividing the rating for target emotion by the sum of other ratings. Of the discussed scoring methods, this method appears to incorporate most of the available information in a single score. On the other hand, the validity of this approach depends on how closely the Likert-scale ratings represent true interval scale (cf. above). Furthermore, it is notable that because the non-target emotion ratings are summed together, the effect of a single confusion is small.

More elaborate methods have been used in transforming ratings into scores. For example, Gosselin *et al* [49] utilized *Signal Detection Theory (SDT)* [58, 59], which differentiates the discriminability of emotion (a signal), usually denoted with d', from the response bias of subjects. The discriminability can be calculated by formula $d' = \Phi^{-1}(H) - \Phi^{-1}(F)$ where H and F denote hit and false alarm rates, and Φ the cumulative distribution function of the standard normal distribution (cf. [58, 59] for this and other SDT calculations). Note that in effect the formula defines how many standard deviations the hit and false alarm rates differ. Gosselin *et al* defined hit rate as the proportion of evaluations where the predicted target emotion was rated higher than other emotions. No definition was given for the false alarm rate, but it was apparently defined as the proportion of evaluations where a considered emotion was rated higher than other emotions from other than intended facial expressions. The use of SDT parameters in emotion judgment studies has similar strengths and problems as the unbiased hit rate: the d' is able to correct subjective biases, but possibly confounds such biases with common confusions between basic expressions.

All studies in the present thesis utilized a rating task where the presence of each basic emotion was evaluated on a Likert scale and the obtained six ratings were converted into a recognition score (as detailed in Chapter 2.2). Bias correction wasn't utilized because it was considered to confound common confusions with subjective biases, as discussed earlier. The used scoring was based on the number of observed confusions, *i.e.* non-target emotions being rated as high as or higher than the target. The used scoring method resembles that used by Kamachi *et al* [55] with two improvements: each observed confusion decreases the recognition score by a constant value and non-target emotions rated higher than target doesn't drop the recognition score to its minimum.

Existing basic expression collections

Available facial expression collections[1] were searched via existing Internet resources [61, 62], Google Internet search engine [63] and a review [64]. The following constraints were set for the collections: i) at least frontal views of the faces should be shown directly; ii) pure facial expression material should be included without simultaneous visual speech or other facial actions; and iii) at least all the six basic expressions should be included. Consequently, some otherwise potential facial collections were omitted. For example, Belfast Naturalistic Database [65] collection contained speech in addition to facial expressions, CMU PIE Database [66] and PICS [67] collections contained only one emotional expression (smiling); and Yale Face Databases [68] (two separate collections) and AR Face Database [69] collection contained only three different emotional expressions (sad, sleepy and surprised; and smile, anger and screaming). Furthermore, some of the expressions in the latter two were not considered emotional (such as sleepy and screaming faces). Seven collections fulfilling the set criteria are presented in Table 1. Most of the collections are available free of charge for the research community by request; however, the TKK collection isn't publicly available at the moment and the availability of CMU (Carnegie-Mellon University Facial Expression Database) collection is usually restricted only to computer vision studies. Importantly, before recent years no collections with dynamic stimuli appear to have been available. At the moment, DaFEX

[1] Facial expression collections have sometimes been called databases. This may in some cases be misleading because the term database hints at the existence of search and retrieval mechanisms for the material (however, see [60] for a counter-example). Here the term collection is used, when possible, instead of database.

(Database of Human Expressions), MMI ("M & M Initiative" Face Database) and UTD (University of Texas Database) are the only available dynamic basic expression collections. EF collection (Pictures of Facial Affects by Ekman and Friesen), the only commercial collection, has been the most widely used facial expression collection in facial emotion studies. It is also the only collection where the published material has been screened from a larger original sample. Most of the collections have been created by asking actors to pose certain FACS action unit configurations on their faces, either after short guidance or after more extensive training. In JAFFE (Japanese Female Facial Expression Database) collection, actors posed emotions freely. Arrangements were such that the actors were able to monitor their faces and take the photographs themselves. In DaFEX collection, professional actors were given short stories depicting certain emotions and asked to pose them by empathizing. The only collection containing spontaneous emotional expressions was the UTD collection, where subjects were videotaped while watching emotion-evoking films. As discussed earlier (cf. Chapter 1.1), some emotions are extremely difficult to evoke, possibly because of display rules, and it is even more difficult to obtain instances of pure basic emotions. Accordingly, the UTD collection contains several instances of happiness, some instances of sadness, disgust and surprise and only a few instances of fear and anger [70]. Ideally, facial expression collections should be evaluated both with FACS coding and by evaluation studies with subjects, although the latter could be claimed to be more important because they confirm that the collected material is actually perceived as intended. CMU and MMI collections have been only FACS coded. It appears that no evaluations have yet been conducted with the UTD collection. TKK collection is the only one with both FACS and subjective evaluations. Surprisingly, it appears that FACS codes are not available for the EF collection published by two of the FACS developers.

Collection	Year	Actors	Type	Elicitation	Screening	Evaluation	Availability
EF [5]	1972	14	Static	FACS (trained amat.)	FACS	Forced-choice (31-147 subjects)[1]	Commercial
JAFFE [57]	1998	10	Static	Posing (amat.)	None	Rating (60 subjects)	By request
CMU [39]	1999	182	Dynamic	FACS (amat.)	None	FACS	By request (limited)
DaFEX [72]	2004	8	Dynamic	Empath. (actors)	None	Forced-choice (80 subjects)	By request
MMI [60]	2005	19	Dynamic	FACS (amat.)	None	FACS	By request
UTD [71]	-	<10[2]	Dynamic	Spont. (amat.)	None	None	By request
TKK (Chapter 2.3)	-	6	Dynamic	FACS (trained actors)	None	FACS and rating (21 subjects)	Not available

Table 1 Currently existing basic expression collections. **Year** depicts the year of publication of the collection's first available description. **Actors** refers to the number of actors with a full set of six basic expressions. **Type** refers to whether the material contains static (pictures) or dynamic (video sequences) stimuli. **Elicitation** denotes how the emotional expressions were created: spontaneously, by posing FACS action unit configurations or by empathizing given emotional states freely; and (in parentheses) whether the actors were amateurs or professional actors and whether they received training before recordings. **Screening** refers to criteria used in selecting the published material from an original larger sample. **Evaluation** denotes how the published set has been evaluated: by FACS-coding or in forced-choice or rating studies. **Availability** refers to whether the collection is available freely by request, liable to charge (commercial) or not available.

[1] Evaluated by several groups; number of evaluators differs between stimuli.

[2] The number of emotional facial expressions ranges from 10 instances or less for anger and fear to 200 for happiness [70].

1.3 Role of spatial frequencies in perceiving faces

Spatial frequencies refer to how frequently a smoothly (more accurately, sinusoidally) alternating pattern of lightness is repeated within a certain spatial distance (cf. [73: pp. 79-84]). For example, in a sine-wave grating with bars alternating smoothly from dark to light and back, spatial frequency refers to the number of repetitions of this pattern within a certain distance. Also more natural objects, such as faces, can be broken into different spatial frequency components where higher spatial frequencies refer to more specific details and lower to more coarse shapes. In relation to faces, spatial frequencies are most often quantified as cycles per face width (c/fw). Using *spatial frequency filtering* methods it is possible to filter certain spatial frequency ranges from images (cf. [74: pp. 201-17]). In *low-pass filtering*, only spatial frequencies up to certain *cut-off frequency* are passed through the filter. Respectively, the lower the cut-off frequency, the more blurred image is obtained. Similarly, only spatial frequencies higher than cut-off frequency are passed in *high-pass filtering*. In *band-pass filtering*, a band of spatial frequencies between lower and upper limits are passed.

Recognition of identity

A study by Fiorentini, Maffei and Sandini [75] indicated worse recognition of identity from spatial frequencies below 5 c/fw than above it, no similar difference at 8 c/fw and better recognition from spatial frequencies below 12 c/fw than above it; however, in practice spatial frequencies higher than 15 c/fw weren't visible. The possible conclusions of this study are limited by the facts that recognition wasn't studied from non-filtered stimuli, specific frequency bands weren't compared with each other and there was an upper threshold for the high spatial frequencies. However, the results suggest that identity recognition was degraded from spatial frequencies below 5 c/fw and that spatial frequencies between 5-12 c/fw may be most important for the recognition of identity. These suggestions are at least partially supported by band-pass filtering studies. Based on three experiments where spatial frequency bands were replaced by noise, distorted or band-pass filtered, Näsänen [76] concluded that identity was recognized best from a

spatial frequency band slightly less than 2 octaves[1] wide and centered somewhere between 8-13 c/fw. Gold, Bennett and Sekuler [77] showed that with 2 octaves wide filters, identity was recognized best at central frequency 6 c/fw (whole frequency band 3-10 c/fw). Identity wasn't recognized at all at 2 c/fw (1-3 c/fw) but was apparently recognized rather well at 25 c/fw (10-40 c/fw). With 1-octave filters, faces were recognized only at central frequency 9 c/fw (6-12 c/fw) by one subject and at 18 c/fw (12-23 c/fw) by another (only two subjects were studied). Identity wasn't recognizable at central frequencies 1, 2 and 4 c/fw (frequency bands 1-2, 2-3 and 3-6 c/fw). Using 1.5-octave filters, Hayes, Morrone and Burr [78] found some improvement over chance at central frequency 6 c/fw (3-9 c/fw) and best identity recognition performance at 13 c/fw and 25 c/fw (7-19 c/fw and 13-37 c/fw). By using low-pass filtering consisting of several superimposed 1 octave wide band-pass filters, Peli and coworkers [79] found better recognition when band-pass filter centered at 8 c/fw (respective band 5-11 c/fw) was included in the superimposed filter than when it wasn't. Note that because lower spatial frequencies were always included, their result highlights the importance of middle in comparison to low spatial frequencies. The authors found also degraded identity recognition when images were distorted by manipulation of 1-octave wide spatial frequency bands centered at 8 c/fw (5-11 c/fw) and 16 c/fw (11-21 c/fw), the former causing higher degradation. Comparison between these studies is made difficult by their methodological differences. Even so, the studies appear to agree on that middle spatial frequencies, possibly centered around 10 c/fw, are more important than low spatial frequencies for the recognition of identity from faces.

A further question is whether low spatial frequencies are of any significance for identity recognition. The results of Gold *et al* [77] showing no recognition of identity from 1-octave spatial frequency bands covering spatial frequencies 1-6 c/fw would suggest that low spatial frequencies are irrelevant. On the contrary, low-pass filtering results by Costen, Parker and Craw [80] show slight but not significantly increasing recognition accuracies and significantly decreasing response times between cutoff

[1] Here *octave* refers to an upper limit twice the lower limit's frequency, two octaves four times the lower limit, *etc*; or in general, $U = 2^c * L$ where c denotes octaves, L the lower limit and U the upper limit. With simple arithmetic, the lower limit can be calculated respectively from the central frequency C by formula $L = 2C / (2^c+1)$. For example, a 2 octaves wide filter centered at C=10 c/fw has lower limit L=4 c/fw and upper limit U=16 c/fw.

frequencies 5, 6, 12 and 23 c/fw. This result suggests that even if low spatial frequencies were less important than middle frequencies, at least low spatial frequencies beginning from 5 c/fw do provide some additional information for the recognition of identity.

Munhall and coworkers [81] studied the recognition of audiovisual speech from band-pass and low-pass filtered faces. Unlike in other discussed studies, dynamic stimuli were used. With 1 octave wide band-pass filters, audiovisual speech was recognized best at central frequency 11 c/fw (7-15 c/fw), although the recognition was almost as good at 6 c/fw (4-7 c/fw). Audiovisual speech was recognized no better than auditory-only speech at central frequency 3 c/fw (2-4 c/fw). With low-pass filtering, on the other hand, some improvement over auditory speech was observed already at cutoff frequency 4 c/fw and recognition accuracy equal to unfiltered faces at 7 c/fw. Interestingly, these central frequencies are lower than those estimated by the identity recognition studies. This could reflect either difference in the spatial frequencies important for the recognition of visual speech *vs.* identity or the use of dynamic instead of static stimuli.

Recognition of emotions

There is relatively little information available on which spatial frequencies are important for the recognition of emotions from facial expressions. It isn't certain whether the results from identity and visual speech studies can be generalized directly on emotion recognition tasks, because these tasks could obviously rely on different spatial frequency bands.

In a study by Nagayama, Yoshida and Toshima [82, 83], subjects made distinctions on whether faces filtered with 1 octave wide band-pass filters were familiar and whether they were smiling or not. Both recognition rates and reaction times were measured. Best performance was observed at central frequency 12 c/fw (8-17 c/fw) for the facial expression and at 25 c/fw (17-33 c/fw) for the familiarity distinction task. However, both expression and identity were recognized rather well at both of these two frequency bands. Interestingly, reaction times for neutral faces were longest and they were mistaken most often as happy at central frequency 6 c/fw (4-8 c/fw) whereas the evaluation of happy faces was not compromised. Although the results were obtained only from neutral and happy faces, this result could suggest that different spatial frequencies are important for

different types of expressions. Schwartz, Bayer and Pelli [84] used a matching task with 21 different emotional facial expressions taken from the same person. Standard basic emotions were included in the posed emotions, along with various other emotional states, but the results weren't analyzed separately [85]. In their study, low- and high-pass noise masks with varying cutoff frequencies were imposed on different locations on the face. As a result, they concluded that performance was most degraded when 8 c/fw noise was added on lower face region near mouth. In a recent fMRI brain imaging study, Vuilleumier and co-workers [86] studied the implicit processing of unfiltered and low- and high-pass filtered fearful faces with respective cutoff frequencies 6 c/fw and 24 c/fw. Their behavioral results from a rating task showed fear to be evaluated as less intense from low-pass filtered than from high-pass filtered and unfiltered faces, indicating degraded recognition of (fearful) emotional facial expressions from low spatial frequencies below 6 c/fw.

Based on these studies, it can be suggested that middle spatial frequencies, possibly around 10 c/fw, are more important than low spatial frequencies for the recognition of emotions from faces, as was the case also with identity recognition. Importantly, the results by Nagayama *et al* [82, 83] with happy and neutral faces give tentative evidence for the fact that different emotional facial expressions would rely on different spatial frequencies.

1.4 Role of motion in perceiving faces

Recognition of identity

Subjects recognize identity better from moving rather than from static displays when presented with dots extracted from original faces of actors [87]. Similarly, identity and sex can be recognized from whole-head and facial movements extracted from human actors and replicated on an animated head showing no original static features from the actors [88]. These results indicate that identity (and sex) can be recognized even from pure motion information.

Other studies have shown that identity is recognized better from moving rather than static faces when the stimuli are shown as negatives (with colors reversed) [89, 90], inverted (turned upside down) or thresholded (presented as monocolor images where

luminances below a threshold are shown as black and those above it as white) [90] or pixelated or blurred [91] but not when shown as originals [89, 90]. It appears that the recognition of identity from static original images is close to ceiling value and isn't facilitated by motion. On the other hand when the facial stimuli have been degraded, as in all of the operations described above, motion compensates for the lack of static information. Respectively, it appears that that static information is the primary modality in the recognition of identity from faces and dynamic information is beneficial only when the static information is insufficient.

Recognition of emotions

In comparison to identity recognition, few studies have been conducted on the role of motion in recognizing emotions from faces. Furthermore, the obtained results have been rather incongruent.

The fact that some neurological patients impaired in recognizing emotions from still images of facial expressions do nevertheless recognize them from video sequences [92] and moving dots [93] suggests that motion information might compensate for the lack of static features also with emotional facial expressions. As with identity recognition, emotions can be recognized from moving dots extracted from the original faces of actors posing different emotions, *i.e.* from pure motion information [87, 94].

Kamachi *et al* [55] compared the effect of stimulus display time on the recognition of emotions from static and dynamic facial expressions, where the latter were created by interpolating artificial movement sequences from neutral to emotional faces. The general difference between static and dynamic stimuli, pooled over different display times, was not significant. However, there was a significant interaction between the type of stimuli (static or dynamic) and display time, evident in that display time had a significant effect with dynamic but not with static faces. Note that because the study concentrated on the effect of display time instead of dynamics, no explicit comparisons were made between dynamic and static stimuli at different display times. It appears, however, that significant effects of dynamics were found. Dynamic stimuli were recognized *worse* than static ones, most notably sad expressions at short displays and angry expressions at long displays. As noted by authors, the results could be affected by the fact that dynamic facial stimuli

starting from neutral face showed the recognizable emotional expression for a shorter time than their static versions. Furthermore, also the artificialness of morphed movement could have affected the results. Morphed movement may differ from biological one because the movement occurs linearly from one facial muscle configuration to another without considering temporal realism or biological constraints on the motion.

Harwood and co-workers [95] studied the recognition of static and moving emotional facial expressions in subjects with and without mental retardation. Both types of subjects recognized sad and angry faces significantly better from dynamic rather than static displays. Wehrle *et al* [6] studied the role of movement in recognizing emotions from a two-dimensional computer animated face and showed that moving facial expressions were recognized better than static ones. It is uncertain whether their result can be generalized to natural faces, because real faces weren't used as control stimuli. In fact, the static displays of animations were recognized poorly, suggesting that the used facial expression model wasn't realistic. Recently, Ambadar and co-workers [96] showed that emotions were recognized better from dynamic rather than static displays of brief movement sequences presented as 3-6 video frames (100-200 ms) from the beginning of full movement sequences. Because the used stimuli were truncated from full video sequences showing transition from neutral face to emotional facial expressions, they were of very slight intensity.

Comparison between the different studies isn't straightforward because of the various types of stimuli used. However, at least the results from Wehrle *et al* [6] and Ambadar *et al* [96] are compatible with the hypothesis based on identity recognition, *i.e.* that dynamics facilitates the recognition of emotions from degraded but not from non-degraded stimuli (cf. above). In the study by Wehrle and co-workers, stimuli were poorly recognized facial animations and in the study by Ambadar and co-workers, emotional facial expressions of extremely slight intensity. However, the results from Kamachi *et al* [55] and Harwood *et al* [95] aren't consistent with the presented hypothesis. In the former study, dynamics apparently decreased the recognition of emotions in some cases. In the latter, emotions were recognized better from dynamic rather than static emotional facial expressions even if the static stimuli were recognized well and not degraded. It is suggested here that the results from Kamachi and co-worker's study are related to its

methodological problems (relatively shorter presentation of emotional displays in dynamic *vs.* static faces and the use of biologically unrealistic movement). On the other hand, no confounding factors are apparent in the study by Harwood and co-workers.

1.5 Asperger syndrome and perception of faces

Although some inconsistencies exist in diagnosing Asperger syndrome (AS), official criteria in ICD-10 [97] and DSM-4 [98] taxonomies (as cited in [7: pp. 13-23]) agree that AS typically involves severe impairments in social interaction, such as impaired use of eye gaze, facial expressions and other nonverbal information of social behavior, and narrow and obsessive interests. General unwillingness towards social communication, preoccupation with parts of objects and rigid routines are also common symptoms of AS. Generally with autistic spectrum disorders, abnormal verbal and nonverbal development is apparent already during the first year of development, including lack of attention towards others and failure to orient to name. The abnormal development becomes more evident during the second and third years of life, including impairments in the use of eye gaze, joint attention and delayed language development [99]. In contrast to other autistic spectrum disorders, AS is thought to involve unimpaired verbal and cognitive skills, including no general delay in childhood language development. No consensus exists on distinguishing AS from high-functioning autism (HFA), most commonly defined as typical autism with moderate or high level of intelligence [7: pp. 23-5]. However, AS has sometimes been characterized as a less severe neurocognitive disorder than HFA with a later onset and more favorable outcome. It has also been suggested that AS is characterized by a general motor clumsiness instead of repetitive movements characteristic to autism, a pedantic speaking style and higher verbal in comparison to non-verbal intelligence [7: pp. 23-5]. AS is comorbid, *i.e.* often occurs, or is confused diagnostically, with at least schizoid and schizotypal personality, attention-deficit hyperactivity and obsessive-compulsive disorders and alexithymia [7: pp. 23-33]. Furthermore, AS has in some studies (cf. [100, 101]) been classified under *social developmental disorders* (SDD) together with other autistic spectrum disorders and socio-emotional processing disorder because of diagnostic overlap between these disorders.

Face processing and recognition of identity

Several studies have suggested general face processing impairments in ASD (see [102, 103] for reviews). Retrospective analyses of home movies have shown that infants later diagnosed with ASD pay little attention towards the faces of others and that the lack of interest for faces is one of the best predictors of a later diagnosis. In comparison to normally developing children, children with ASD typically perform worse in discriminating and recognizing faces, fail to show better memory performance for faces *vs.* other stimuli, concentrate on atypical facial features and utilize featural instead of configural processing of faces (see below).

Theory of weak central coherence, an influential neuro-cognitive theory of autism, claims that a general tendency for concentrating on local instead of global features of objects is characteristic for ASD [8]. Rigid routines and preoccupation with object parts, as well as outstanding skills on restricted areas such as calendar calculation or musical competence sometimes observed with autistic individuals [8], could reflect such bias. Enhanced processing of local and impaired processing of global features has been suggested by several studies (see [104] for a review). For example, individuals with ASD have been found to perform better than neurotypical individuals in detecting local targets and embedded figures from visual stimuli. On the other hand, they have been found to fail in perceiving impossible visual figures, which requires perceptual integration of parts, and in using grouping heuristics to understand inter-element relationships.

Weak central coherence theory is relevant in a discussion of face processing because faces have been suggested as a category of complex natural stimuli requiring sophisticated configural processing. It has been suggested that three configural processing types are related to the perception of faces [105]: sensitivity to first-order relations characterizing the general configuration of faces (two eyes located above nose *etc*); sensitivity to second-order relations, *i.e.* specific distances among facial features, considered important for identity recognition; and holistic processing interconnecting facial features into a unified percept, evident for example in that differences between facial features are recognized worse from full faces than when shown in isolation. On the other hand, featural processing is related to perceiving the shape, color or luminance of individual facial features, such as eyes or mouth. Inversion of faces, *i.e.* turning faces

upside-down, has been shown to disrupt all configural processing types but have little effect on featural processing of faces (see [105] for a review).

As with non-face objects, individuals with ASD have been suggested to have enhanced featural and inferior configural processing of faces in comparison to neurotypical individuals (see [102, 104, 106] for reviews), although the latter finding has recently been debated [106]. Some studies have shown that individuals with ASD pay more attention on lower face features than neurotypical individuals while watching interacting people, hinting at general peculiarities in observing faces. Enhanced featural processing is supported by the finding that unlike neurotypical individuals, individuals with ASD show priming effect for individual facial features when recognizing faces. Inferior configural processing of faces has been supported by the finding that individuals with ASD show less impaired recognition of identity from inverted faces than neurotypical individuals (as reviewed in [102, 104]). However, some recent studies have shown typical inversion effect in ASD and found other evidence for a typical configural processing of faces (as reviewed in [106]). For example, children with ASD have been found to be prone to “Thatcher illusion”, where the inversion of mouth and eye areas is less apparent in inverted than upright faces. Individuals with ASD have also been found to be better in recognizing isolated facial features when previously encoded in a face context. In conclusion, the earlier studies on face perception give good support for the hypothesis that individuals with ASD tend to use local information in processing faces; however, whether this also implies degraded configural processing is less certain.

In a recent study by Deruelle and co-workers [107], 11 children with autism or AS between ages 4-13 years were compared with typically developing children in an identity-matching task with low- (below 12 c/fw) and high-pass (above 36 c/fw) filtered faces. Children with ASD were found to perform significantly better with high-pass in comparison to low-pass filtered faces whereas an opposite result was observed with typically developing children. Although the authors didn’t explicitly test whether the performance with high- and low-pass filtered faces differed between the groups, the groups’ opposite results suggests that this effect was significant. Because featural information is obviously evident in the high spatial frequencies and configural information in the low spatial frequencies, the result is congruent with the hypothesis of

degraded configural processing in individuals with ASD. Interestingly, it was found that the children with ASD performed the better in matching low-pass filtered faces the older they were. This suggests that during childhood and adolescence, individuals with ASD learn to compensate an initial deficit in configural processing to some extent.

Only a few studies have studied whether adult individuals with AS disorders have *prosopagnosia, i.e.* whether they are impaired in recognizing identity from faces (see [101, 108] for reviews). Barton *et al* [108] compared identity recognition in subjects with SDD disorders (including AS), patients with prosopagnosia and neurotypical controls. As a result, they found that 8 out of 24 subjects with SDD performed equally to controls whereas 16 were impaired in recognizing identity, although most of them performing better than typical prosopagnosic patients. The results didn't differ between subjects with AS and other SDD diagnoses. This study indicates that although face recognition impairments are common in individuals with AS (and other SDD disorders), such impairment isn't always present. In a continuation study by Hefter and co-workers [101], the relation between identity and emotional facial expression recognition was compared between subjects with SDD. Their results suggested no correlation between identity and expression recognition skills. Furthermore, no difference was observed in the recognition of emotional expressions between subjects with and without identity recognition impairment. This result is compatible with patient and brain imaging studies that have indicated a dissociation between identity and emotion recognition from faces [109].

Recognition of emotions

Skill of the individuals with ASD in recognizing emotions from facial expressions is of specific importance for this thesis. The results from earlier studies are conflicting. Several studies on basic emotions have suggested that autistic children and young adolescents are impaired in processing emotional facial expressions in comparison to typically developing children and/or children with other developmental disorders [107, 110-113]. In these studies, autists have shown worse accuracy in comparison to controls in finding the odd facial expression out of a set of facial expressions and naming emotional facial expressions [110], matching faces on the basis of emotional expression [107, 111], matching emotional labels with human and (curiously) orangutan and canine

faces justifiably depicting emotional states [112], and equal accuracy but longer response times in naming emotional facial expressions [113]. These results appear to be due to a general face processing deficit, as some of these studies using several face processing tasks found the autistic subjects to be less accurate than non-autistic subjects also in recognizing identity [110], gaze direction and visual phonemes from faces [107] (a non-significant difference in the recognition of identity was found at two studies [107, 111]; however, this could have been due to the small subject samples of 10 and 11 autistic children).

Several other studies have failed to replicate the finding that autistic children (and young adolescents) [114-116] were impaired at recognizing basic emotions from faces. Such studies have shown non-significant differences between subjects with ASD and neurotypical controls in sorting pictures of emotional facial expressions [114], matching emotional facial expression video sequences with pictures [115] or matching verbal labels with emotional facial expressions [116]. The last study [116] found that children and adolescents with AS recognized emotional facial expressions paired with mismatching emotional words worse than controls, suggesting either a higher-level emotion processing deficit or a general deficit in suppressing responses to verbal stimuli (for the latter interpretation, see [106]). Baron-Cohen and co-workers [117] found a significant difference between autistic and non-autistic children in the recognition of surprised, but not happy or sad expressions. Because a sorting task between the emotional stimuli was used, their results could be taken as evidence for a deficit in recognizing basic emotions from faces. However, the authors claimed that autistic children were impaired specifically in the recognition of surprise because unlike happiness and sadness, it required attaching a belief to the observed person ("being surprised *about* something"). This appears a far-flung conclusion because a plausible and simpler explanation for their result would be that the surprised facial expressions used in their study were the most difficult expressions to discriminate among the used stimuli. A similar study using facial expressions of all six basic emotions selected from the EF collection [5] found no specific impairment with autistic children in the recognition of surprise [114]. However, the deficit in processing the beliefs of others in ASD suggested by Baron-Cohen *et al* [117]

has been supported by later studies from the same first author with more refined distinction between basic and complex mental states (see below).

The bulk of research on emotional facial expression recognition in ASD has concentrated on infancy, childhood and adolescence with only a few studies with adult participants. Studies with adult individuals with AS and HFA have consistently suggested typical recognition of basic emotions but a deficit in making more complex social judgments in comparison to neurotypical controls [118-122]. The consistency of these results suggests that even if impaired in recognizing basic emotions in childhood, high-functioning autistic individuals learn compensatory strategies during their later development. Adult individuals with HFA have been shown to give abnormally high ratings on the approachability and trustworthiness of faces in comparison to controls [118]. Similarly, studies on adult individuals with HFA and AS have suggested impaired recognition of "complex mental states" requiring the attachment of beliefs or intentions to the observed person (such as threat, regret, astonishment, worry and distrust [123]) from faces, especially from eyes [119-121]. A recent study has suggested similar impairment in recognizing complex emotions both from faces and voice [122]. The failure of adult individuals with ASD to recognize the alleged complex mental states from faces and eyes suggests a deficit in interpreting the mental states of others. According to an established neuro-cognitive theory of autism, such "theory of mind" (*aka* mentalizing) deficit is most fundamental for ASD [8].

1.6 Overview of studies

The purpose of two first studies is to evaluate research material used in later studies, *i.e.* TKK collection containing video sequences of basic expressions (Chapter 2.3.1, cf. Table 1) and facial animations created with "TKK talking head" (Chapter 2.3.2), a three-dimensional animation model of a talking and emotionally expressive person. The recognition of basic expressions is compared between TKK and the more standard EF [5] collections (**study I**[1]), and the TKK talking head is compared to two other talking heads and one human actor (**study II**). Third and fourth studies address the main hypothesis, *i.e.* that dynamics improves the recognition of basic expressions only from degraded stimuli, directly. The recognition of basic expressions is compared between static and dynamic stimuli created with TKK talking head, with a further comparison with posed facial expressions selected from CK collection [39] (**study III**); and the recognition of basic emotions is studied with static and moving facial expressions blurred at different levels with low-pass filtering (**study IV**). In the last study (**study V**), the foregoing study is replicated in a revised form with persons with AS to evaluate whether autistic spectrum disorders involve untypical recognition of basic emotions from moving and degraded facial expressions. Research hypotheses specific to the conducted studies are stated in their descriptions.

[1] Numbering of studies reflects conceptual rather than temporal order. The studies were conducted in the following order: study III, study II, study I, study IV and study V.

2 RESEARCH METHODS

This chapter describes research methods common to all studies. Specific details are described separately for each study.

2.1 Procedure

In all studies, stimuli were either pictures or video sequences of facial expressions depicting the six basic emotions. Stimuli were presented with Presentation software (versions 0.53 to 9.51[1]) [124] in a randomized order. The subjects were sitting approximately 80 cm from the monitor. Stimuli were shown on a 19" (18" viewable area) monitor with a resolution of 1024×768 pixels. The size of stimuli varied in different experiments. The subjects' task was to evaluate how well each of the six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) applied to the presented facial expression; in studies I-III, subjects were also asked to evaluate how natural the facial expressions appeared (either six or seven evaluations for each stimulus). The subjects were told that each facial expression might contain none, one or several of the presented emotional labels to avoid biasing them on categorical evaluations. When the naturalness evaluation was used, subjects were explained that naturalness referred to whether a human actor actually experienced any emotions or whether a computer animation resembled a genuine expression of emotion. The different questions were shown in random order for each evaluated stimulus and answers were given, using a keyboard, on a 7-step Likert scale ranging from total disagreement (1) to uncertainty (4) and to total agreement (7). Similar emotion judgment studies have typically used 4- [42], 5- [6, 56, 92, 125], 7- [5, 49, 55, 126] or 10-step [19] Likert scales. Note that in these studies the task has been evaluating the intensity of expressed emotions, typically ranging from none to severe, whereas in the present thesis the task involves evaluating the suitability of different emotion labels on facial expressions. The subjects were instructed to answer quickly on the basis of their first impression; however, no limits were set on the response time.

[1] With each study, the latest available version was used. The used version didn't affect the experimental procedure.

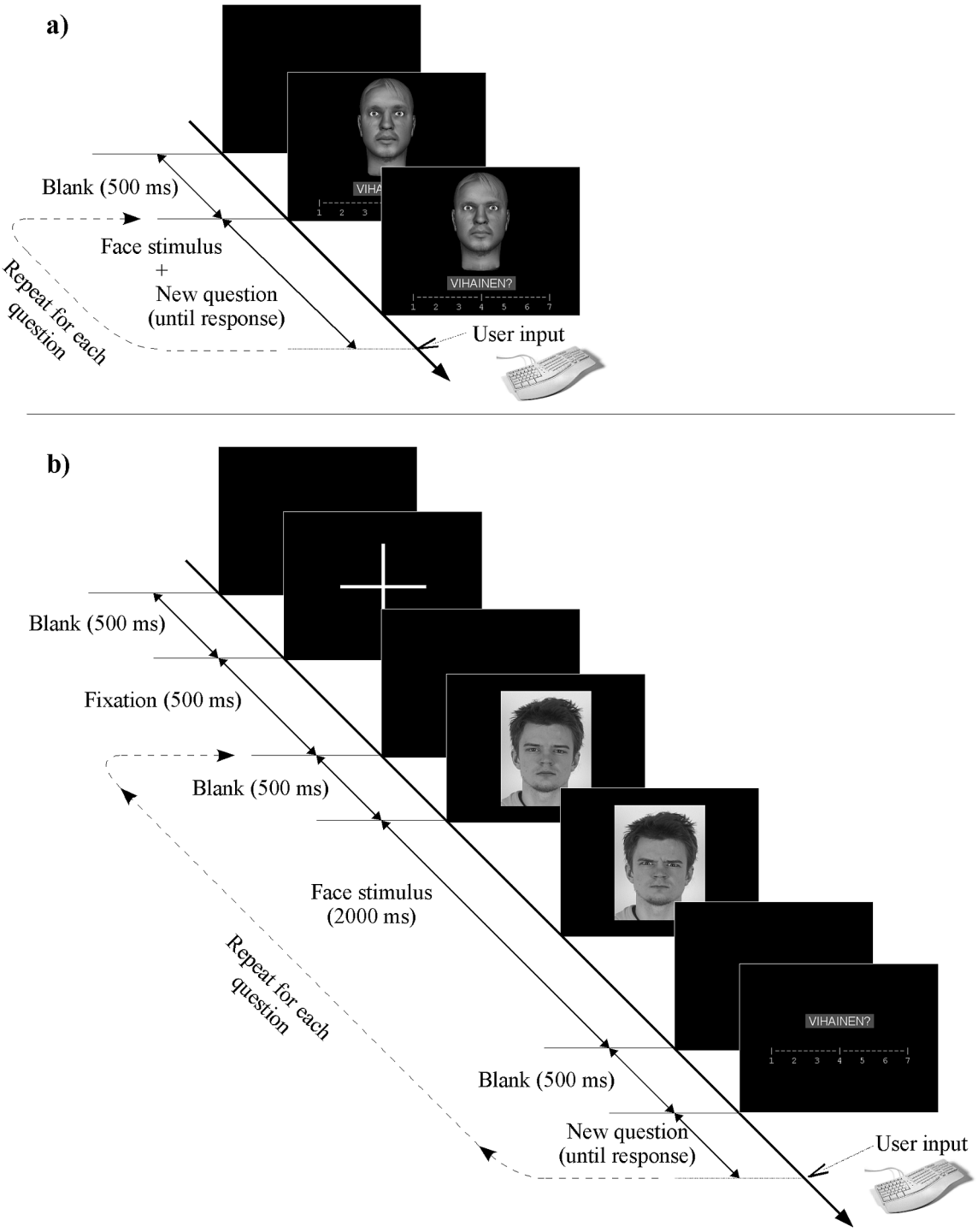

Figure 1 Presentation of one evaluated stimulus, *i.e.* picture or video sequence depicting a basic expression, in different studies. a) In studies II and III, questions were presented below evaluated stimulus that remained visible until response was received for each presented question. Pictures remained stationary and video sequences were played in a loop with prolonged presentation (500 ms) of the last frame. b) In studies I, IV and V, stimulus was repeated once with a fixed duration before each question. Pictures remained stationary and video sequences were played once with last frame kept visible until the whole video sequence had been presented for the intended duration.

The stimulus presentation is illustrated in Figure 1. In studies II and III, stimuli and questions were presented together and each stimulus remained visible until responses had been received for all questions. This allowed for different viewing times of stimuli by different subjects. In study II based on repeated-measures design, this was unlikely to affect the main results because exactly the same stimuli were evaluated by all subjects. In mixed-design study III, the effects of viewing times were specifically tested for in the analysis. In studies I, IV and V, each stimulus was repeated once for a fixed duration (2 s) before each presented question. A procedure where each stimulus was presented only once before the presented questions was tested in pilot studies; however, this task appeared too difficult. In studies I, IV and V a separate fixation mark (white cross in studies I and IV, and gray square in study V) was used to indicate the change of evaluated stimulus after all questions for a stimulus had been answered (cf. Figure 1). With each study, the actual experiment was preceded by a training session. Studies II-IV involved one rest break and longer studies I and V two rest breaks.

In studies I, III and V the subjects were asked to fill a TAS-20F (the 20-item Toronto Alexithymia Scale; translated in Finnish) self-report questionnaire [127] measuring alexithymic personality trait [128]. The TAS-20 overall score is a sum of three distinct alexithymia factors: difficulties in recognizing emotions (factor 1), difficulties in describing emotions (factor 2) and a tendency for external thinking (factor 3). Results from a conclusive alexithymia evaluation study in Finnish population [129] were used as reference values. Alexithymia level was tested in study I to evaluate whether it had any effects on the evaluation of basic expressions and was used in studies III and V to ensure no differences between subject groups. In study II alexithymia level wasn't evaluated because only a single subject group was used. Alexithymia level wasn't evaluated in study IV. Although this was somewhat inconsistent with studies III and V, no significant alexithymia differences should be expected because the subject groups resembled those used in study III.

2.2 Statistical analyses

Statistical analyses are conducted with t-tests and analyses of variance (ANOVAs) with a significance level α=0.05, Bonferroni-corrected with formula α^c=0.05/*n* for *n* multiple comparisons when appropriate. In some cases exploratory analyses are conducted where the significance level isn't corrected for multiple comparisons because inflating the number of missed effects (type II error) was considered worse than inflating the number of mistakenly accepted effects (type I error). Most statistical analyses are conducted with Statistica software (version 5.5) [130].

Error correction

During the experiments, several subjects reported making at least one erroneous response. The errors were often such where a subject gave an opposite answer than intended (for example, a key response 1 instead of 7). A conservative error correction procedure was used, in which only such obvious outliers were corrected. A given rating was considered an outlier if it was at least five points lower/higher than the lowest 10%/90%[1] of all subject's ratings (*i.e.* 10^{th}/90^{th} percentiles [52]). Such outliers were replaced with subjects' median ratings. For example, if all ratings except the lowest 10% were 6 or 7, ratings 1 would have been replaced with the median. It wasn't possible to simply remove the outliers because in the analysis of missing data would have required the existence of full data sets from at least two subjects [130], which was very unlikely since all subjects evaluated several actors and several facial expressions in each study. The error correction results are reported individually for each study.

Scoring methods

The six emotional ratings for each evaluated stimulus are converted into a *recognition score S* with formula

$$S = \frac{n_{u<i} - n_{u\geq i}}{n} = 1 - \tfrac{2}{5} * n_{u\geq i}$$

[1] This would have required that *n*-1 were a multiplier of 10 (where *n* denotes the number of subjects). In practice the lower percent threshold was defined as the multiple of 100 / (*n*-1) *closest* to 10, or lower of the multiples if two of them were equally close to 10. The higher percent threshold was defined as the lower threshold subtracted from 100. For example, if the number of subjects was 26, the percentage thresholds would have been 8% (2∗100/25 %) and 92% (100 − 2∗100/25 %).

where n refers to the number of unintended emotional ratings (other than the target emotion; always equal to 5), $n_{u \geq i}$ to the number of unintended emotional ratings equal to or higher than the intended target emotion and $n_{u<i}$ to the number of unintended emotional ratings lower than the target (equal to $n_{u<i} = n - n_{u \geq i} = 5 - n_{u \geq i}$). The resulting score is a value in range [-1…1] reflecting how distinguishable the target emotion is from the other emotions. For example, if the target emotion were rated higher than all other emotions, a recognition score of 1 would be given. Each confusion, *i.e.* a non-target emotion receiving at least as high rating as the target, decreases the recognition score by 0.4 ($^2/_5$). If all other emotions were rated equal to or higher than the target emotion, the score would be -1. *Chance recognition level*, where ratings would be given at random to all emotions, would produce an expected number of 2.5 confusions and a mean recognition score of 0. Note that the chance level may be exceeded even if the target emotion is confused systematically with some other emotions. It is suggested here that for an emotional facial expression to be recognized distinctively, its recognition score should exceed an *ambiguous recognition level* of 0.6 (confusion with one unintended emotion) significantly. In a large subject sample, this criterion is fulfilled when few subjects make no confusions and the remaining at most one confusion. More stringent criteria were tested in practice, but they were found to be too strict.

Instead of analyzing the non-interval scale naturalness ratings for a certain stimulus directly, the ratings are converted into a *naturalness rate* reflecting how large proportion of subjects considers the stimulus natural to any extent (naturalness rating>4, the "uncertain" rating).

Confusion analysis

The transformation from several emotion ratings to one recognition score (cf. Chapter 1.2) represents a compromise between information richness and the ease of interpretation. The recognition scores are more informative than simple hit rates and easier to interpret than the original ratings. However, the ease of interpretation comes at the cost of losing some of the original information. Unlike original ratings, recognition scores obviously fail to indicate what kinds of specific confusions occurred in the evaluation.

For being able to evaluate the recognition of individual emotions, each of the six ratings for an evaluated stimulus is converted into *emotion recognition rate*[1], defined as the proportion of subjects that have given the emotion a rating that is both their highest rating among the six emotional ratings for a stimulus (note that the highest rating can be shared by several emotions) **and** higher than the uncertain rating (4). The latter constraint is used to avoid misattributing high recognition *rates* in a case where low ratings, such as the uncertain rating, have been given to several or all of the emotions.

When analyzing emotion recognition *rates*, the null hypothesis is that no emotions have received ratings above the uncertain rating (4), producing zero *rates* for all emotions. Individual emotions are considered to be recognized from an evaluated stimulus when their recognition *rates* exceed zero significantly.

It should be noted that whereas recognition scores consider only the rating differences between each non-target and the target emotion, emotion recognition *rates* consider each emotion's rating in the context of all other ratings. The used approach was considered appropriate because it makes it possible to evaluate how often both the target and non-target emotions were perceived distinctively.

Response time analysis

For the analysis of response times, means were calculated over all ratings of each evaluated stimulus.

2.3 Research stimuli

Studies I-V used static and dynamic emotional facial expressions both from real human faces and computer animations. No video sequence collections were available when the first of these studies were conducted (cf. Chapter 1.2). A new facial expression collection containing video sequences of six basic expressions and some of their blends was recorded in the Laboratory of Computational Engineering, TKK ("TKK collection"). The expressions were posed by professional actors. Simple emotional facial expression

[1] Naturalness and emotion recognition *rates* are always denoted in italics, when appropriate, to distinguish them clearly from *ratings* and recognition *scores*.

animations were created with a facial animation model developed at the Laboratory of Computational Engineering ("TKK talking head") [131].

Emotional facial expression prototypes (Appendix B) were defined with FACS [31] by a certified FACS coder (JK) both to be posed by the actors for the TKK collection and to be modeled on the TKK talking head. The prototypes were based on the basic emotion theory containing six basic emotions, two of their blends and a non-emotional facial expression; note however that all of the studies I-V used only facial expressions of basic emotions. The prototypes were intended rather as typical examples of basic expressions than definite models of them, because no unequivocal and undisputable FACS prototypes exist for them (cf. Chapter 1.1). The devised prototypes were designed on the basis of tentative prototype suggestions given by the authors of FACS (Appendix B), a publication from two FACS authors containing verbal and illustrated descriptions of emotional facial expressions [24], and a comprehensive facial expression guide for artists containing anatomical, verbal and illustrated descriptions [132]. Blends of basic expressions were based on the basic expression prototypes and additional sources [24, 133]. All basic expression prototypes also resembled the FACS authors' original prototype suggestions (Appendix B), except the addition of AU7 (lid tightener) into the prototypes of fear and sadness on the basis of [24].

The suitability of the devised prototypes was confirmed by checking their emotional interpretations from the FACSAID dictionary [35] (cf. Appendix B). For each prototype, a practically identical action unit combination with the intended interpretation was found. The differences were negligible and apparently related to the following coding conventions adopted in FACSAID: inclusion of head- and eye position coding (AUs higher than 50); laterality coding (coding related to uni- vs. bilateral facial actions); ignoring mouth opening caused by action unit activations, such as AU25 (lips part) caused by AU20 (horizontal lip stretch) or AU26 (jaw drop); or coding action units with overlapping effects together, such as AU4 (brow lowerer) with AU9 (nose wrinkling) where the latter causes some brow lowering resembling that of the former. The latter two conventions are contrary to the guidelines given in FACS instructions [31].

Action units were classified further into primary and secondary action units on the basis of [3, 24, 31, 33: pp. 173-4] (Appendix B). *Primary action units* refer to the most important emotional facial actions without whom the intended emotional messages would either change or become less intense. *Secondary action units* refer to actions that could either be caused by the primary action units or co-occur with them without altering the intended emotional messages significantly. Note that the suggested full prototypes were in most cases combinations of primary and secondary action units.

2.3.1 TKK basic expression collection

Figure 2 A sample set of pictures (in gray-scale, originals in color) from the TKK collection. The initials of actors from left to right are SP, KH, TV, NR, MR and ME. Intended emotions from left to right are anger, disgust, fear, happiness, sadness and surprise.

Six actor students and graduated actors (3 men and 3 women, age range 23-32 years) were recruited from the Theatre Academy of Finland to pose emotional facial expressions. One basic expression from each actor is shown in Figure 2.

The TKK collection was intended both for psychophysical and computer vision studies (for a computer vision study based on the TKK collection not reported in this thesis, see [134]). Such facial expression stimuli were required that would by themselves depict single basic emotions as clearly as possible in the absence of contextual or other information. The computer vision studies also required similar expressions from all actors. The method of posing FACS-based facial configurations was selected to produce distinctive and similar emotional facial expressions. In general, successfully posed facial expressions resemble those of genuine emotions, although possibly in an exaggerated form [4] (cf. Chapter 1). The use of FACS prototypes ensured that all actors posed relatively similar facial actions. A disadvantage of FACS-based posing is that the facial expressions may appear unnatural; especially the dynamics of facial muscle activations

may be artificial and the intensity of facial actions may differ between different sides of the face [14, 135]. It was acknowledged that it might not be possible to fully solve this problem. However, to improve the naturalness of posed expressions, the actors were encouraged to empathize the required emotions. The actors underwent a long training period to familiarize themselves with the required facial configurations.

The actors read written instructions designed by a certified FACS coder and practiced the required facial configurations (Appendix B) individually for 5–10 hours. The instructions included earlier facial expression poses from two FACS coders (JK and VK) as illustrations. A practice recording session was held, at the middle of the practice period, where the actors were able to inspect their facial expressions carefully from replays and to give and receive feedback prior to the actual recording session. Feedback from the actors was used to adjust some of the emotion prototypes. Most notably, an open-mouthed happiness variant was added which appeared more natural than the original closed-mouthed prototype.

To improve the applicability of the collection for computer vision studies, recordings were made both with and without additional markers attached on the face. With the former, nine markers were placed on emotionally salient locations which were found difficult to track by computer vision algorithms. To reduce head movements, the actors were asked to keep as still as possible while posing the emotions on their faces. The actors were asked to pose each emotion 5–10 times, each pose beginning from a neutral face and ending to the emotional facial expression. A FACS coder (JK) selected those recordings which were estimated most similar to the intended emotion prototypes to be included in the collection. Because of the small number of actors posing emotions, all selected basic expressions were included in the collection without screening. The duration of the final video sequences was 1.2±0.3 s (mean±s.d.; range 0.7-1.8 s). Please note that facial expressions depicting other than basic emotions or showing additional markers weren't evaluated in any of the studies in this thesis.

The used recording setup included a digital camcorder (Sony DSR-PD100AP) and two professional photographing lamps (Elinchrom Scanlite 1000). The video sequences were recorded 25 frames per second (fps) with horizontal interlacing and a resolution of 576×720 pixels. Some interlacing artifacts were clearly evident in recordings with quick

facial actions, especially with facial expressions of surprise. To remove these artifacts, the original video sequences were deinterlaced by selecting every odd horizontal line from the original frames and resizing the obtained half-frames vertically to one half of the original height. The final results were deinterlaced video sequences with a resolution of 288×360 pixels.

2.3.2 TKK talking head

Background

Talking heads [136] are three-dimensional facial animation models of a talking person including a visual phoneme articulation model synchronized with auditory speech synthesis, sometimes including also an animation model for facial expressions. Talking heads are typically used to synthesize audiovisual speech from text with additional command tags for controlling facial expressions. Also more sophisticated automatic non-verbal behaviors have been included for emphasizing spoken text and for enlivening emotional states (see [137] for an example). Talking heads have been utilized in various applications such as multi-modal computer interfaces [138], low-bandwidth teleconferencing [139], speech therapy [140] and entertainment [141]. Talking heads can be utilized also in audiovisual (cf. [142]) and emotional facial expression related neurocognitive research. Ideally, facial animations generated with talking heads have some advantages for research use: the animations can be created easily, are fully controllable and contain no unwanted movements. Three-dimensionality makes changing viewpoints and head positions easy.

The traditional facial animation techniques can be classified at least into four fundamental methods [143: pp. 105-148]. In *interpolation* or key-framing method, full facial expressions are devised by animators and the rest of the animation is interpolated between these key frames. In *performance-driven* animation, the animation is driven by real human actions, measured for example by laser- or video based motion tracking. Interpolated and performance-driven animations are ideal for producing good-quality animations, such as those required in movies and other entertainment, but creating new animations or modifying existing ones is laborious. In *muscle-based* animation, the characteristics of facial muscles and/or other facial tissues are simulated to produce facial

expressions, as in [144-146]. A potential drawback is that such simulations tend to be computationally heavy. In *parameterization* or parametric *method*, a small set of control parameters is used to transform the facial surface. Parametric model was first devised by Parke [143], and has been used since in several facial animation systems, e.g. [147, 148]. A parametric animation model is computationally light, but it can be used for approximating realistic facial expressions. A drawback of this method is that subtle but important facial tissue changes, such as skin wrinkling or bulging, are not modeled. Furthermore, the final result depends on what kinds of transformations are used and on the skill of the animator setting up the parameters.

Because FACS system [31] is intended for describing all visually distinguishable changes on the face caused by underlying facial muscle activity, it can be used as a basis for facial animation. Respectively, FACS has been utilized in several parametric [136] and muscle-based animation models [144, 145]. More recently, a systematic definition for synthetic facial animations resembling FACS has been given in an international MPEG-4 [139] multimedia standard. In MPEG-4, facial movements are represented with several facial animation parameters (FAPs) resembling FACS action units closely.

TKK talking head implementation

Only the emotional facial expression modeling and overall implementation of TKK talking head[1] are described here, for a more detailed technical description please refer to [131]. As usually is the case in three-dimensional computer animation, the TKK talking head is based on polygonal modeling where the head shape is defined as a mesh of polygons defined by interconnected vertices, *i.e.* points in a three-dimensional space [150]. The TKK talking head uses a facial mesh from the University of Washington [151] with additional eyes and teeth, modified eye openings and improved mouth region (Figure 3). The talking head is

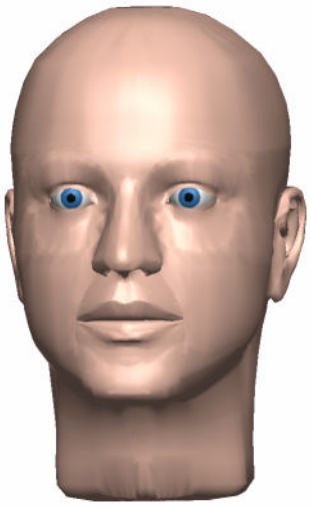

Figure 3 TKK talking head with rendered facial mesh.

[1] Note that the present talking head is an independent and completely remade version of an earlier Finnish-speaking talking head also developed at the Laboratory of Computational Engineering, TKK [149].

capable of producing articulatory lip movements synchronized with auditory speech synthesis, and facial expression animation. The articulatory movements for vowels and consonants, measured from real three-dimensional data, can be combined for producing visual speech from text.

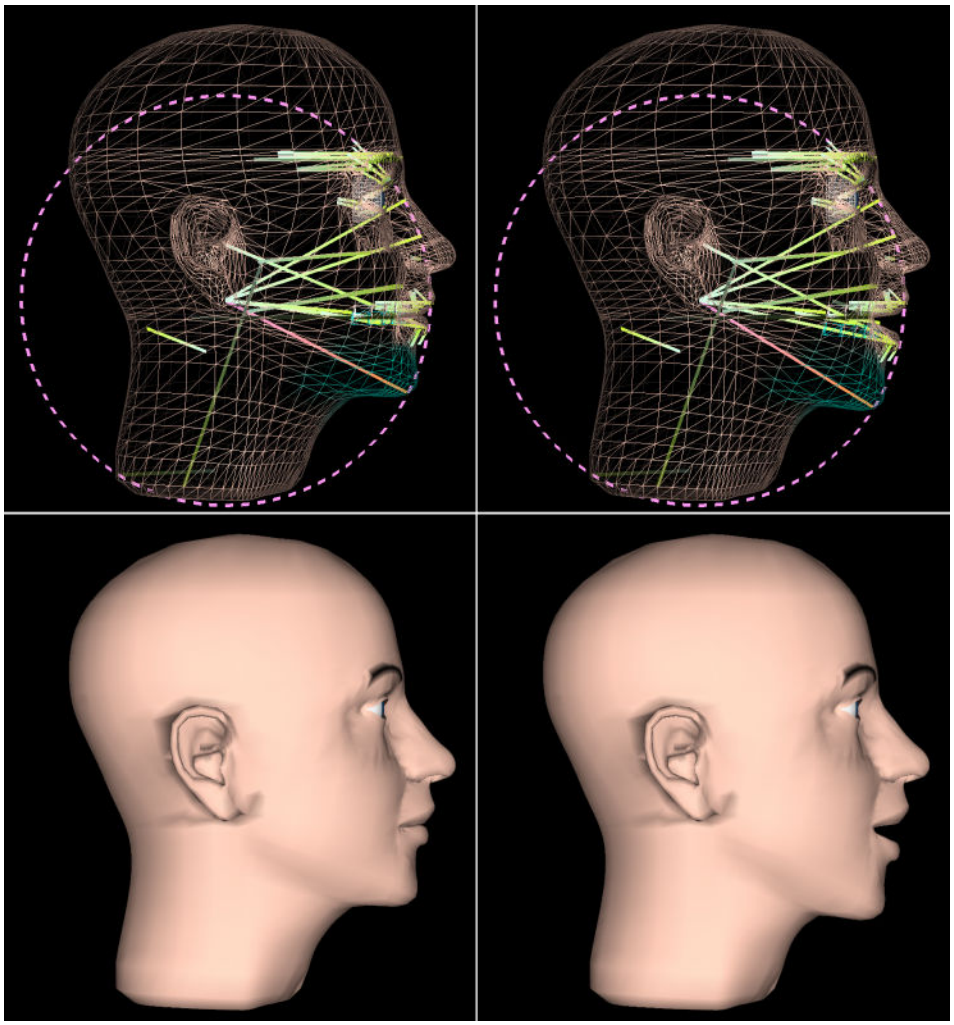

Figure 4 The effects of jaw opening parameter on the TKK talking head. Upper: Wire-frame model of the talking head, showing the facial mesh and all parameters (in light green) from one side. Jaw opening parameter (in violet) has been selected, and its area of influence on the mesh (in turquoise) and its rotational plane (circle in turquoise) are shown. Lower: The talking head with rendered surface. The neutral state is shown on the left and the changes caused by the parameter on the right.

TKK talking head implements Parke's parametric model with an additional parameter hierarchy controlling the effects of overlapping parameters, as described in [131]. The facial mesh is manipulated with geometric transformations. Although any transformations could be added, the current implementation uses only rotational transformations. Respectively, the parameters can be thought of as rotational deformators with a center point, radius and an influence region on the facial mesh defined as a set of vertices. The influence weight of a parameter is defined separately for each of its vertices. When the value of a parameter is changed, the positions of influenced vertices are transformed along a circular plane defined by the parameter, as illustrated in Figure 4 (above). An additional feature of the TKK talking head is texture mapping front and side photographs of a real person's face on the facial mesh (Figure 5). It is also possible to reshape the talking head's head shape to match that of the photographed person [149]; however, the deformation parameters aren't automatically adjusted to suit the new facial mesh.

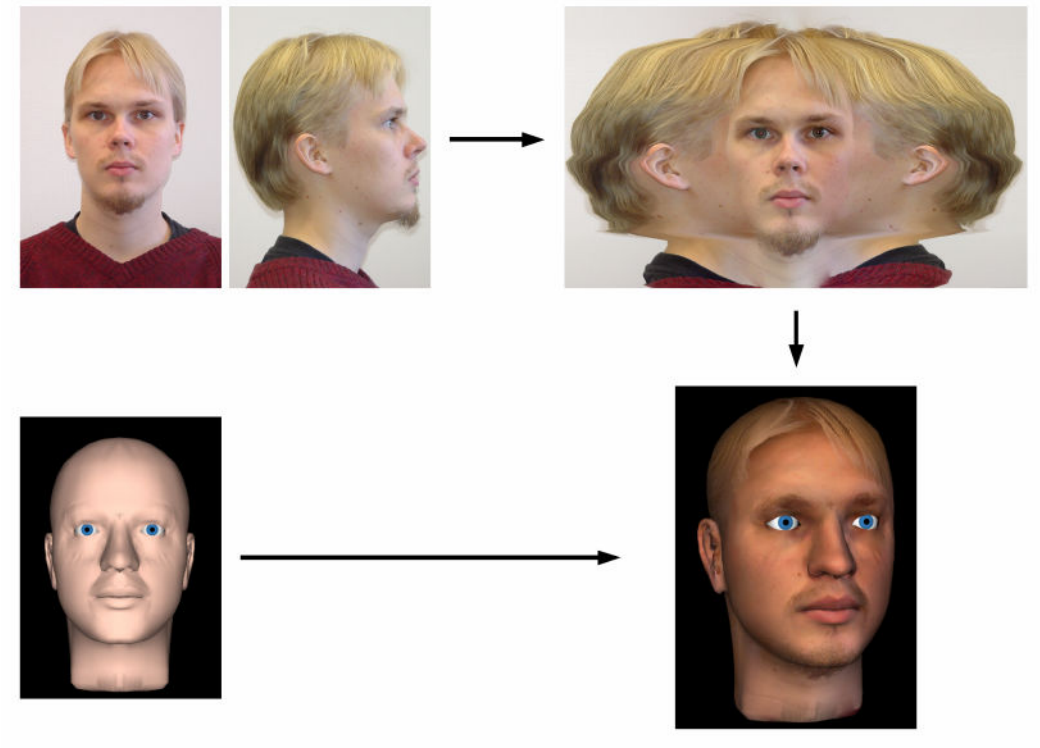

Figure 5 Texture mapping face photographs on the TKK talking head.

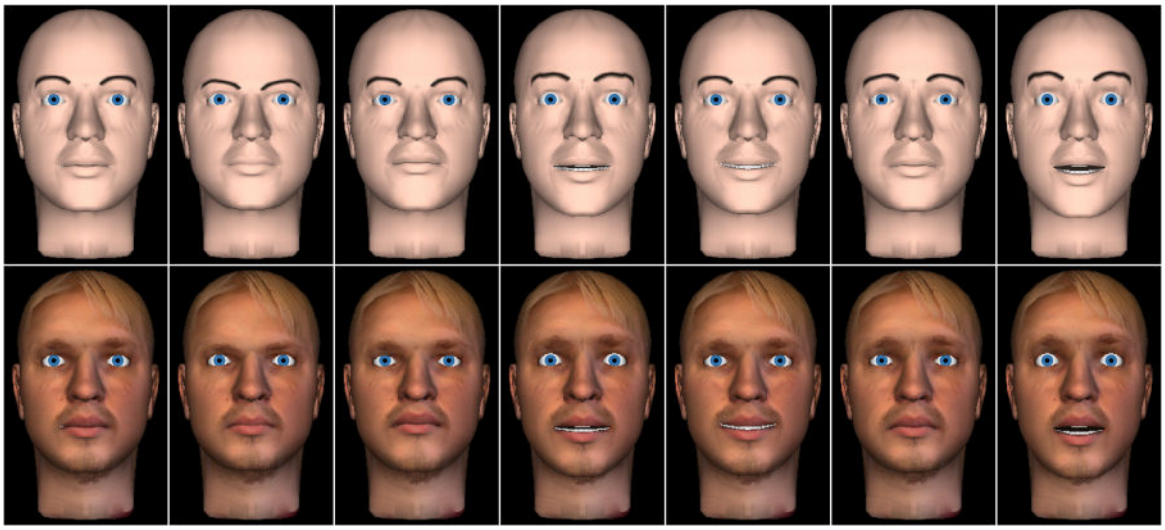

Figure 6 Basic expressions of non-textured (upper pictures) and textured (lower pictures) versions of the TKK talking head. Intended facial expressions are from left to right: neutral, anger, disgust, fear, happiness, sadness and surprise.

A FACS coder (JK) set up the parameters for 20 different action units (Appendix D) manually. These action units were used to implement the FACS based emotional facial expression prototypes (Appendix B.2). The resulting synthetic TKK talking head facial expressions, both with and without additional facial texture, are shown in Figure 6 (above). The use of FACS based modeling of facial expressions, the use of parametric animation model and especially the constraint on rotational transformations set some severe constraints on the implementation of basic expression animations. Because FACS is mainly concerned with static facial changes, it doesn't provide information on the temporal dynamics of emotional facial expressions. Consequently, the facial expression movements were reduced to purely linear transitions from neutral to emotional faces. As usually is the case with parametric models, subtle featural changes such as skin wrinkling and bulging were not modeled. Because a single rotational transformation can produce movement only along a circular plane, the manual definition of movements along more complex surfaces was practically impossible. As a consequence, only the beginning of activations could be modeled with some action units, limiting the facial expressions to slight intensities. Orbital activations (constricting activations around eyes or mouth) could not be modeled directly with rotational transformations and had to be approximated by several parameters. For example, a slight lowering of outer brows was added to the main action of raising cheeks in AU6 (cheek raiser). Because the implemented emotional

facial expressions were of low intensity and lacked important featural changes, some of the emotions weren't evident on static pictures of facial expressions. However, the emotions seemed more apparent when the simple linear movement was added. The effect of movement is studied explicitly in study III (Chapter 4.1).

3 EVALUATION OF RESEARCH STIMULI

Before conducting facial expression studies, careful evaluation of stimuli is necessary for ensuring that they are perceived as intended. Objective FACS coding and subjective emotion judgment studies are commonly used in evaluation (Chapter 1.2). This chapter describes FACS evaluation for the TKK collection and evaluation studies for the TKK collection and TKK talking head.

3.1 TKK collection FACS analysis

FACS coding procedure

All basic expressions in TKK collection were FACS coded by one certified FACS coder (JK), and part of them (24 of the 36 stimuli) comparison coded by another certified coder (VK). The results are presented in Appendix C. Inter-observer agreement for an evaluated stimulus was defined as two times the number action units agreed by both coders divided by the number of all coded action units, a measure recommended by FACS authors [33: p. 17]. The resulting agreement rate was a value between [0…1]. Note that miscellaneous action units [31], such as jaw clenching and nostril dilation, were considered unimportant and ignored. The mean inter-observer agreement (0.85±0.03; mean±s.e.m.) was significantly above the mean agreement (0.76) between independent coders obtained in an evaluation study by FACS authors (t_{23}=3.44, $p<0.03$, α=0.05)[1] [33: p. 19]. Inter-observer agreement on intensity was evaluated with the further constraint that the intensity evaluations had to be within one step from each other in the 5-step intensity scale used in FACS. This criterion was used instead of exact agreement because guidelines for the 5-step intensity coding have been evaluated to be subjective [152]. The mean inter-observer agreement (0.72±0.04) didn't differ significantly from the foregoing agreement found by FACS authors. The first coder used the comparison coder's evaluations to revise his original FACS coding, resulting in a slight increment in mean agreement on facial actions (0.86±0.03) and intensity (0.76±0.04). Further analyses were based only on the first FACS coder's evaluations.

[1] This analysis is only suggestive, however, because no statistical deviation parameters were given in the study.

Analysis of FACS codes

The coded FACS action units were classified further into those that were primary, secondary and extra (neither primary nor secondary) to the intended basic expression prototypes (Appendix B). An overview of the results (Appendix C) is shown in Table 2. Primary action units were lacking in 10 out of 42 (24%) emotional facial expression poses; however, of these only one (fearful emotion posed by actor SP) lacked more than one primary action unit (cf. Appendix C). Various extra action unit combinations were observed in 23 (55%) of all poses.

Posed emotion	Ok	Missing	Extra	Both
Anger	2		4	
Disgust			6	
Fear		4		2
Happin. (opened)	6			
Happin. (closed)			5	1
Sadness		1	3	2
Surprise	6			
All	14	5	18	5

Table 2 TKK collection stimuli classified on the basis of basic expression prototypes. The columns refer to number of stimuli containing all primary and no extra action units ("Ok"), all primary and some extra action units ("Extra"), lacking some primary and no extra action units ("Missing") and both lacking primary and containing extra action units ("Both"). See text for further details.

The obtained results obviously vary between posed emotions. Happy (opened-mouth variant) and surprised facial expressions were exactly as intended, as they contained all primary and no extra action units. All angry and disgusted and virtually all of the closed-mouth happiness expressions contained all primary actions; however, most of them contained also some extra actions. One closed-mouth happiness expression (by actor MR) contained AU13 (sharp lip corner puller) resembling that of the intended primary action unit AU12 (lip corner puller). With this one exception, only fearful and sad facial expressions lacked primary action units. Three out of six sad facial expressions lacked one of the facial actions considered primary for sadness. With one exception, all sad facial expressions contained also extra actions. All fearful facial expressions lacked primary action units and two of them contained extra actions. With four actors the lacking primary action was AU7 (lid tightener) and with one actor AU5 (upper lid raiser).

Discussion

The analysis of FACS codes shows that with one exception, all of the posed emotional facial expressions contained most of the facial actions considered primary for basic expressions. Because none of these facial expressions lacked more than one primary action, they can be expected to depict the intended emotional facial expression prototypes rather well. On the other hand, several other factors may influence the recognition of emotions from these facial expressions. Most importantly, a majority of the facial expressions were found to contain extra actions possibly distorting the intended emotional messages. Furthermore, this analysis didn't consider the intensity of facial actions and the timing or dynamics of facial actions. A further evaluation with subjects is obviously necessary to confirm that the collected stimuli are perceived as intended.

In general, the results suggest that posing action unit combinations exactly and without additional facial actions is an extremely difficult task even after considerable training. Of the used basic expression prototypes, happiness and surprise appeared most easy and fear and surprise the most difficult to pose. Furthermore, it appears that it was easier to pose the opened-mouth than the closed-mouth variant of happiness.

3.2 TKK collection evaluation study (study I)

The recognition and naturalness of basic facial expressions in the TKK collection were evaluated in a study, where a representative sample of pictures selected from the EF collection [5] served as comparison stimuli. The main goal was to compare how well the emotional expressions were recognized from TKK collection and how natural they appeared in comparison to the EF stimuli. The EF collection is a good reference in this kind of evaluation because of its very wide use.

Recognition results were expected to be best for happy and worst for fearful and disgusted facial expressions because of known confusions between basic expressions (cf. Chapter 1.1). The used recognition scoring (Chapter 2.2) made it possible to compare results both to chance level and to ambiguous recognition level where the target emotion was confused with at least one emotion. The results of individual actors' expressions were compared to chance level to exclude any obviously faulty evaluations from further analysis. The results of different basic expressions, averaged over all actors, were

compared to the ambiguous recognition level for evaluating whether all basic expressions were in general recognized unambiguously. The hypothesis was that both of these comparisons would exceed the used comparison level significantly.

The effects of subject-related factors on recognition results were also evaluated. Results were compared between male and female subjects evaluating male and female actors. This comparison was motivated by a study indicating male subjects to be worse than female subjects in recognizing disgust from the faces of female actors [153]. No prior hypotheses were made for this evaluation. The effect of an alexithymic [128] personality trait, defined as having difficulties in expressing and experiencing emotion, was evaluated. Earlier studies have indicated alexithymia to be associated with a reduced ability in recognizing facial emotional stimuli [56, 154]. Consequently, the hypothesis was that higher alexithymia level would be inversely correlated with the recognition of emotions.

Methods

Research methods follow those described in Chapter 2 with the changes and additions defined here.

Stimuli

Stimuli were 72 pictures of facial emotions selected from 12 actors each posing a set of six basic expressions.

Six actors (3 male and 3 female; initials JJ, PE, WF, C, NR and PF) were selected from the EF collection, based on an original forced-choice evaluation study [5]. Only actors posing all six basic expressions were used. If several pictures of the same emotion existed for an actor, the best recognized picture was selected. However, opened-mouth happiness variants were always selected if both opened- and closed-mouth variants were available from an actor. This was done because whereas opened-mouth happiness expressions were available from all actors in the EF collection, closed-mouth variants were not. As similar as possible male-female pairs were selected in terms of the mean recognition percentages. The mean recognition rate calculated over the three same-sex actors was 91% for both male and female actors. Mean recognition rates were between 89-94% for different actors and between 74-97% for different posed emotions.

All six actors (KH, ME, MR, NR, SP and TV) of the TKK collection (Chapter 2.3.1) were used in evaluation (see Figure 2). To match the stimuli selected from EF collection, all of the selected facial expressions stimuli depicted basic emotions, didn't contain additional markers and were presented as pictures instead of video sequences. Similarly, only opened-mouth happiness variants were selected. Pictures were obtained by selecting last frames from the original video sequences. The recognition of emotions from dynamic and static facial expressions is compared explicitly in study IV (Chapter 4.1).

All original EF pictures contained number labels identifying the actor and posed emotion. These labels were erased with an image editing program. The TKK pictures were edited to match the appearance of the EF stimuli by cropping the pictures to include only the face area and ears. All pictures were resized to a resolution of 200×300 pixels (7 cm × 10 cm on the screen) and converted into gray scale.

Subjects

Twenty-one subjects (11 male and 10 female) with an age range 19–47 years (mean age±s.e.m. = 26±2 years) participated in the experiment. Seventeen subjects were students from the Helsinki University of Technology (TKK) and four from the Open University of the University of Helsinki (OUH). No results were analyzed separately between these groups because of the small number of subjects in the latter group. All subjects were native speakers of Finnish and had either normal or corrected-to-normal vision.

The subjects were asked to fill TAS-20F self-report questionnaire [127] measuring alexithymic personality trait [128], defined as having difficulties in expressing and experiencing emotions. The TAS-20F overall or factorial scores did not differ statistically from the reference values of Finnish population [129]. The scores didn't differ significantly between male and female subjects.

Error correction

With emotion ratings, the average number of error corrections per subject was 0.5. No more than 3 corrections were made for any subject. Only one naturalness rating of one subject was corrected.

Results

Recognition accuracy

Mean recognition scores for the TKK and EF collections, pooled over actors, are shown in Figure 7. The scores were analyzed with a 2 (collection) × 6 (expression) repeated-measures ANOVA. Only the main effect for expression was significant ($F_{5, 100}$=16.95, p<0.0001). Post-hoc comparison with Newman-Keuls test showed that fearful facial expressions (0.57±0.06; mean score ± s.e.m.) were recognized worse than all other emotional facial expressions. Happiness (0.97±0.01) was recognized significantly better than all other emotional expressions except surprise (0.93±0.02). Differences in the recognition of disgust, anger and sadness (mean 0.80±0.03) were not significant. The main effect of collection or the interaction between collection and expression were not significant. Planned comparisons for different emotional expressions suggested slightly better recognition of disgust from TKK rather than EF collection (0.86±0.03 *vs.* 0.79±0.03; $F_{1,20}$=5.15, p<0.04); however, this difference didn't reach corrected significance level (α^c=0.05/6=0.008).

To evaluate differences between individual TKK actors, a 6 (actor) × 6 (expression) repeated-measures ANOVA analysis was conducted. The main effect of actor was found to be significant ($F_{5,100}$=2.38, p<0.045). Post-hoc analysis with Newman-Keuls test showed the only significant difference to be that between actors MR (0.85±0.03) and SP (0.77±0.04). However, when results for fearful expressions were ignored, contrast analysis between these actors failed to reach significance (recalculated mean scores 0.93±0.02 *vs.* 0.86±0.03). Fearful expression of actor SP was found to be unsuccessful in a later analysis (see below).

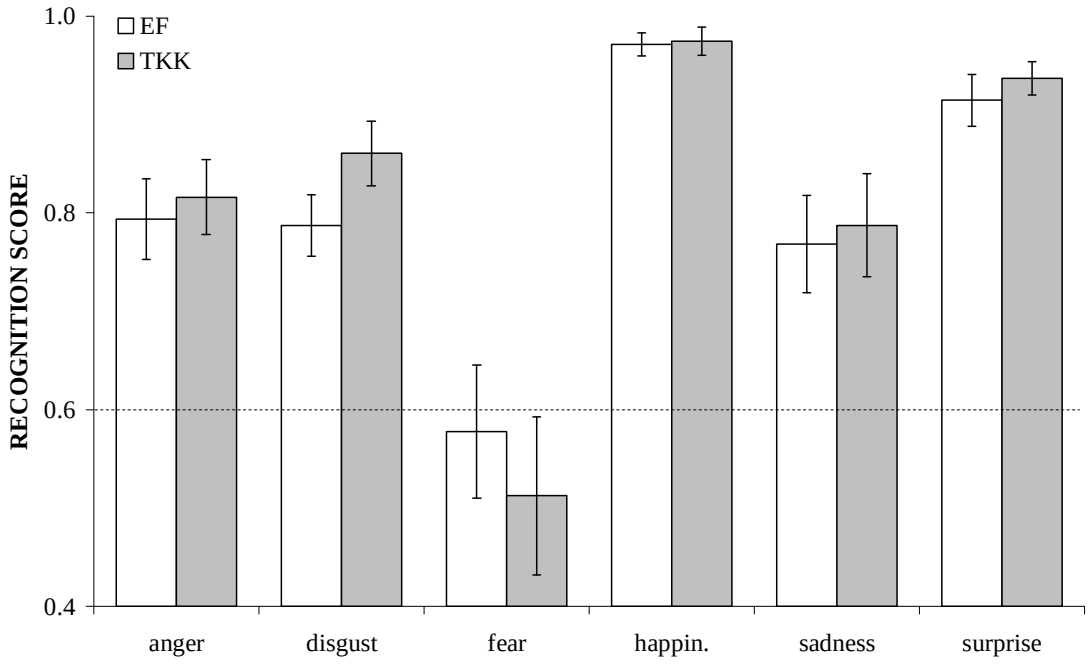

Figure 7 Mean (± sem) recognition scores for different facial expressions from EF and TKK collections. No differences between the collections were significant ($p>0.008$). Ambiguous recognition level (recognition score 0.6; confusion with one emotion) is indicated with a dashed line.

Naturalness

The *naturalness rates* for different emotions were averaged over actors in a collection (Figure 8) and analyzed with a 2 (collection) × 6 (expression) repeated-measures ANOVA. In general, the EF stimuli were considered more natural than the TKK stimuli (0.70±0.06 *vs*. 0.58±0.06; $F_{1,20}=28.51$, $p<0.0001$). A significant main effect of expression ($F_{5,100}=6.57$, $p<0.0001$) indicated also differences between the posed emotions. Post-hoc comparison with a Newman-Keuls test showed higher naturalness evaluations for happiness (naturalness rate 0.83±0.04) than for other emotional facial expressions (mean 0.60±0.06) and no significant differences between the latter. The interaction between collection and expression was not significant, suggesting roughly equal difference between the collections for all emotional expressions.

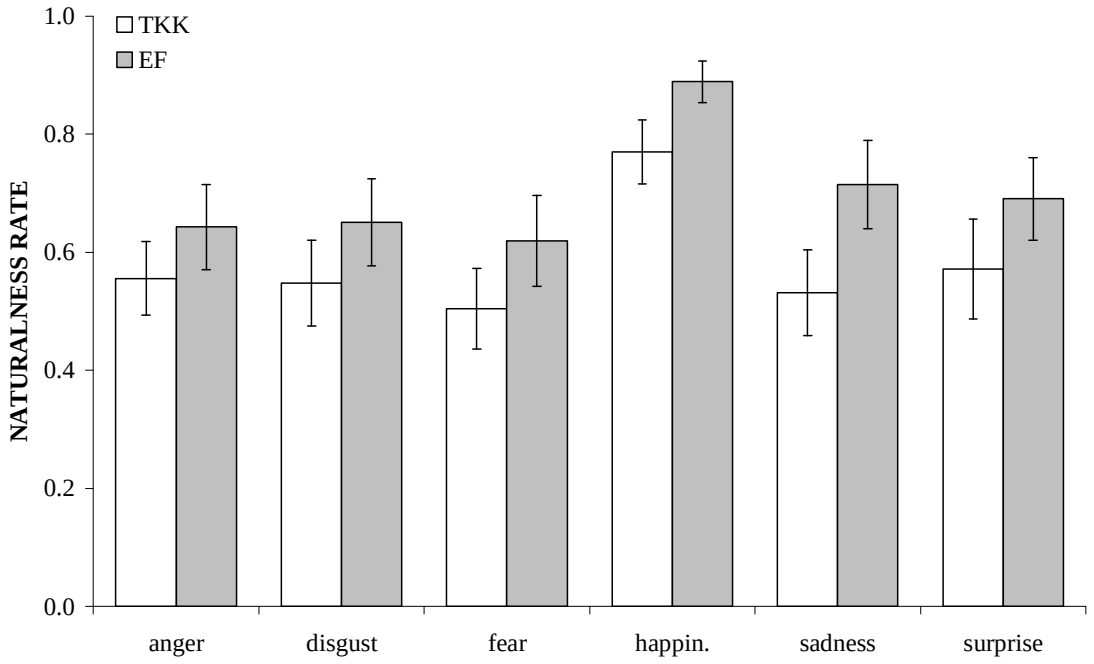

Figure 8 Mean (± sem) *naturalness rates* for different facial expressions from EF and TKK collections. Please refer to text for the description of significant differences.

Differences between individual TKK actors were evaluated with a 6 (actor) × 6 (expression) repeated-measures ANOVA. The results showed a significant main effect of actor ($F_{5,100}$=7.36, p<0.0001). Post-hoc comparison with a Newman-Keuls test showed that actors KH (0.67±0.07) and MR (0.71±0.07) were evaluated natural significantly more often than the remaining actors (mean *rate* 0.52±0.05).

Confusions

Recognition scores for individual actors' emotional facial expressions were compared to chance level threshold (score 0) with multiple one-tailed t-tests. The fearful facial expression of TKK actor SP failed to exceed chance level significantly (recognition score −0.03±0.15), indicating clearly that this expression wasn't perceived as intended. Consequently, recognition score results for fearful expressions were removed from all analyses unless stated otherwise. All other emotional facial expressions exceeded chance with extremely strict significance level correction (α^c=0.05/72=0.0007). Analysis of *emotion recognition rates* indicated that the only significantly (α^c=0.05/6=0.008) recognized emotion from SP fear was surprise (recognition *rate* 0.57±0.11; t_{20}=5.16, p<0.0001), while the recognition of fear (0.19±0.09) failed to reach corrected

significance level. Further comparison with one-tailed t-test confirmed that surprise was recognized significantly more often than fear from SP fearful expression (t_{20}=2.36, p<0.015).

The mean recognition scores for basic expressions were compared to ambiguous recognition level (recognition score 0.6) with multiple one-tailed t-tests (cf. Figure 7). Because the interaction between collections and expressions wasn't significant, the results were averaged over all actors regardless of collection. All other expressions except fear (score 0.55±0.07) exceeded the ambiguous recognition level significantly (α^C=0.05/6).

For evaluating the most common confusions characteristic to different basic expressions, *emotion recognition rates* of different emotions were averaged over all actors. Most common confusions for different basic expressions, *i.e.* non-target emotions with the highest mean recognition *rates*, are presented in Table 3. All non-target emotions not included in the table received mean recognition *rates* smaller than 0.10. Recognition *rates* for the target emotions were compared to those of the most commonly confused emotions with one-tailed t-tests. Because the mean recognition scores of all basic expressions except fear exceeded the ambiguous recognition level, correction for multiple comparisons wasn't considered necessary. The results showed that target emotion was recognized significantly (α=0.05) more often than the most common confusion with all basic expressions except fear, with whom the most common confusion was surprise. With fearful facial expressions, 95% confidence interval for the difference between emotion recognition *rates* of fear and surprise ranged from −0.11 to 0.36[1]. It should be noted that even the upper confidence limit, *i.e.* 100−36%=64% of evaluators **not** recognizing fear more often than surprise, reflects a considerable confusion. An exploratory analysis on all actors without multiple comparison correction showed that the recognition *rate* of fear exceeded that of surprise only with EF collection actors CC (mean difference 0.52±0.16; t_{20}=3.20, p<0.003) and JJ (0.43±0.16; t_{20}=2.63, p<0.008) and TKK collection actor TV (0.52±0.11; t_{20}=4.69, p<0.0001).

[1] Calculated as $\bar{x} \pm t_{0.05}(20) * s = 0.13 \pm 2.09 * 0.11$

Basic expression	Target	Confusion	Difference
Fear	0.66±0.05	Surprise: 0.53±0.07	0.13±0.11
Disgust	0.88±0.02	Anger: 0.33±0.04	0.55±0.05 *
Sadness	0.85±0.03	Disgust: 0.11±0.03	0.74±0.06 *
Anger	0.88±0.03	Disgust: 0.10±0.02	0.78±0.04 *
Surprise	0.96±0.01	Fear: 0.13±0.03	0.83±0.04 *
Happiness	1.00	Surprise: 0.03±0.01	0.96±0.01 *

Table 3 Mean *emotion recognition rates* (± s.e.m.) for target emotions, the most commonly confused emotions and their mean differences for different basic expressions. Significant ($p<0.008$) differences between target and the most commonly confused emotion are marked with an asterisk ('*').

Other factors

For analyzing whether the sex of subjects had any influence on the recognition results, a new analysis was conducted with a 2 (sex of subject) × 2 (sex of actor) × 6 (expression) mixed-design ANOVA. As a result, the main effect of "sex of subject" and all of its interactions were non-significant. However, a significant main effect of "sex of actor" was observed ($F_{1,19}=16.00$, $p<0.0008$), evident in that female actors received higher mean recognition scores than male actors (0.84±0.02 *vs.* 0.76±0.03).

Correlations were tested between recognition scores, response times, and TAS-20F overall and factorial scores with Spearman's correlation tests (Table 4). The first TAS-20F factor was considered to be of specific importance because it is directly related to difficulties in recognizing emotions [127]. A significant ($\alpha^c=0.05/5=0.01$) negative correlation was observed between response times and recognition scores ($r=-0.73$, $t_{19}=-6.25$, $p<0.0001$). A significant positive correlation was observed between TAS-20F factor-1 scores and response times ($r=0.45$, $t_{19}=2.94$, $p<0.006$). A slight but not significant correlation between TAS-20F overall scores and response times was also observed ($r=0.31$, $p=1.91$, $p=0.06$). The correlations between the second or third alexithymia factor and recognition scores or response times failed to reach significance with significance level $\alpha=0.05$.

	Recognition score	Response time
Recognition score	-	-0.73 *
TAS-20F	-0.13	+0.31
TAS-20F F1	-0.25	+0.45 *
TAS-20F F2	-0.22	+0.26
TAS-20F F3	+0.16	-0.01

Table 4 Correlations between recognition scores, response times, and TAS-20F overall and factorial scores. Significant ($p<0.01$) correlations are marked with an asterisk ('*').

Discussion

The TKK and EF collections were compared by evaluating their overall recognition and naturalness measurements. No significant difference was found between the recognition of emotions from TKK and EF collections, suggesting good overall recognition of emotions from the TKK stimuli. As anticipated, EF facial expressions were considered more natural than TKK facial expressions. This result possibly reflected a more tedious selection process for the EF collection, for which "hundreds of photographs were studied over a period of several years" [5]. Intended basic emotions were recognized above chance level from all TKK stimuli except the fearful expression of actor SP. When this unsuccessful expression was left out, no significant overall differences were observed between different actors in the TKK collection. On the other hand, two actors (KH and MR) received higher overall naturalness rates than the remaining TKK actors.

As expected, fearful expressions received the worst and happiness the best scores. On the other hand, the results for disgust didn't differ significantly from the remaining facial expressions. Analysis of recognition scores indicated that all basic expressions except fear exceeded ambiguous recognition level (confusion with one emotion) significantly. The most common confusions, recognizing surprise from fearful and anger from disgusted facial expressions, and the virtual absence of any confusions with happiness replicated the results from earlier studies [3, 4, 16, 19, 22: p. 266, 38]. Notably, the results suggested that most of the fearful facial expressions were actually perceived as blends of fear and surprise. The confusion between fear and surprise is probably explained by physical similarities between them. The used prototypes for fear and surprise (Appendix B) resembled each other on three facial areas: brows (AU1+2+4 for fear *vs.* AU1+2 for surprise), eyes (AU5+7 *vs.* AU5) and mouth (AU20+25+27 *vs.* AU25+26/27). The brow and eye area activations for the fearful facial expression prototype are especially difficult to pose and failures may produce facial activations prototypical to surprise. This was evident with the fearful expression of TKK actor SP, which was perceived more surprised than fearful. The FACS coding for this facial expression (Appendix C) confirms that its brow (AU1D+2D) and eye (AU5B) activations were closer to the surprised rather than the fearful prototype.

The perception of fearful facial expression as a blend of fear and surprise differs clearly from the results of two earlier rating studies with EF stimuli [5, 19]. In the study of Ekman and co-workers [5], over 95% of subjects rated target emotions at least one point higher than all other emotions on a 7-step scale. In another study [19], surprise received consistently the second highest rating for fearful facial expressions, but it was almost always rated lower than fear and equal ratings between fear and surprise *never* occurred [37: p. 274]. The differences between the present and the earlier studies can't be due to a non-representative selection of EF stimuli, because the used stimuli were recognized well in an original forced-choice evaluation study. Two plausible explanations for the observed result are suggested. First of all, because all subjects were Finnish, the results could reflect cultural differences in the evaluation of emotions. For example in a rating study by Matsumoto and Ekman [155], Japanese subjects made a similar confusion between fear and surprise as was observed in the present study whereas American subjects recognized fear unambiguously. Alternatively, the results could reflect differences between the rating scales used in the present and earlier rating studies (cf. Chapter 2.1). Unlike earlier studies using intensity evaluation ranging from none to strong emotion, the present study used agreement evaluation with a fixed middle-point (the "uncertain" rating). For example, the present scale included only three positive response choices (ratings higher than uncertainty) whereas a 7-step intensity scale beginning from nil intensity in a study by Ekman *et al* [19] contained 6 positive choices. Respectively, the used scale might have failed to discriminate subtle intensity differences between fear and surprise. However, even if this were the case, differences in the recognition of fear and surprise from fearful faces were obviously small.

No significant effects related to the sex of subjects were found in this study. However, the recognition of emotions was found to be more distinctive from female than from male actors. This result could reflect either that female actors are in general more proficient in posing emotions or that the perception of sex interacts with the perception of emotional facial expressions. Which one of these alternatives is true is out of the scope of this thesis.

A negative correlation between response times and recognition scores was observed. This result could indicate that the level of subjects' emotion recognition skills was inversely related to their required evaluation times. Alternatively, it could be that the more evaluation time was used, the more unintended emotional features were perceived. The fact that response times were correlated also with the first TAS-20 alexithymia factor, related specifically to difficulties in recognizing emotions, gives support to the former explanation. On the other hand, it is true that correlation between the first alexithymia factor and the actual recognition scores failed to reach significance. The fact that emotion recognition difficulties were related to response times but not to recognition scores could be due to the rather easy task of evaluating exaggerated posed emotions.

3.3 Talking heads comparison study (study II)

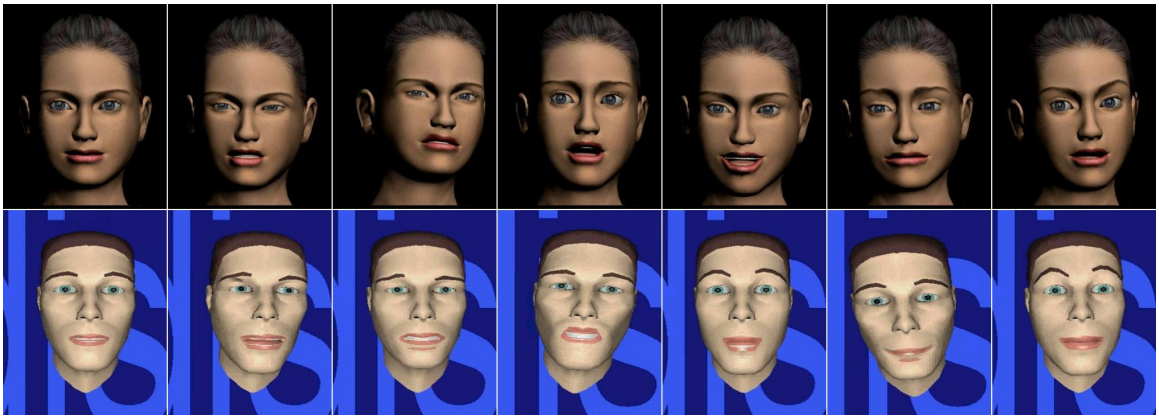

Figure 9 Emotional facial expression animations by Miralab (up) and Image Coding Group (down). The intended facial expressions from left to right are: neutral, anger, disgust, fear, happiness, sadness and surprise.

Recognition and naturalness evaluations of basic expressions from TKK talking head (Chapter 2.3) were compared to those of two other animated talking heads with a full set of six basic emotions (Figure 9) provided by MIRALab, University of Geneva, Switzerland [156] ("ML") and by Image Coding Group, Linköping University [157] whose animations utilized facial animation engine from Department of Communications, Computer and Systems Science, University of Genova, Italy [158] ("ICG/DIST"). Video sequences obtained from a real person were used as control stimuli.

All of the evaluated talking heads were based on parametric animation (cf. Chapter 2.3.2). On the other hand, high-level facial expressions of each head were based on a different model: the expression prototypes for TKK talking head were derived from literature (Chapter 2.3, Appendix B), ML facial animations were designed by an artist on the basis of pictures and video clips taken from real actors [156], and the ICG/DIST facial animations were created by motion-tracking MPEG-4 [139] compliant markers on the faces of actors [157, 159]. The talking heads differ also on some other details. The ML and ICG/DIST models were based on MPEG-4 standard whereas the TKK facial animation model was based on FACS. This should, however, have a negligible effect on the quality of facial animations as MPEG-4 standard is capable of animating most of the FACS action units [31, 139]. ML and ICG/DIST facial animations contained rigid whole-head movements, those of TKK did not. ICG/DIST facial movements were driven by motion-tracked real facial movements and the ML facial animations contained customized linear transitions from neutral to emotional facial displays [156]. In contrast to ICG/DIST talking head, both TKK and ML heads contained a realistic facial texture superimposed on the general face model. In ICG/DIST head, a low-resolution facial model was used without additional texture for creating a facial expression model with low computational demands.

The emotional facial expressions of ML talking head haven't been evaluated earlier. With ICG/DIST facial animations, the recognition of emotions from the human actors and their motion-tracking based facial animations has been evaluated in a study with more than 100 subjects [157, 159]. The results showed that the ICG/DIST facial animations were recognized clearly worse from the animations than from the original faces from which they were motion-tracked from.

The hypothesis was that emotional facial expressions would be recognized worse from all the evaluated talking heads than from the human actor. This is a justified expectation because all of the talking heads aimed for realistic unexaggerated facial expressions on one hand but on the other hand lacked important fine details on the face such as skin wrinkling. The overall recognition level of all animated talking heads was compared to chance and ambiguous recognition levels. The hypothesis was that the recognition would exceed chance level with all of the evaluated talking heads. It was expected that TKK

emotional facial animations would be recognized better and evaluated more natural than ICG/DIST animations because of the lower resolution and the lack of facial texture in the latter. Because of their low intensity, TKK animations were expected to be recognized worse than those of ML head. In addition, the latter was expected to be evaluated as more natural because of the rigid head movement model, more complex facial expression dynamics and the overall emphasis on visual appearance.

Methods

Research methods follow those described in Chapter 2 with the changes and additions defined here.

Stimuli

Stimuli were 24 short (1-2 s) video sequences of six basic expressions. Four sets of stimuli were prepared, each containing the six expressions from one source. One set of expressions contained items from a real human actor (initials KH) selected from the TKK collection (Chapter 2.3); other three sets contained animated talking head facial expression provided by TKK (Chapter 2.3), ML [156] and ICG/DIST [157, 158]. All stimuli were resized to a resolution of 288×360 pixels (10 cm ×12 cm on the screen) and scaled so that the face was approximately of the same size in each stimuli set. Each video sequence showed the expression from neutral face to an emotional apex.

Subjects

Subjects were 12 employees at the Laboratory of Computational Engineering, Helsinki University of Technology, who participated in the experiment as volunteers. All subjects had either normal or corrected-to-normal vision.

Error correction

As a result of error correction, no modifications were made on any emotional or naturalness ratings.

Results

Recognition accuracy

Mean recognition scores for the evaluated stimulus sets, pooled over facial expressions, are shown in Figure 10. A repeated-measures ANOVA indicated that the differences between stimulus sets were significant ($F_{3,33}$ = 43.82, p<0.0001). Post-hoc analysis with Newman-Keuls test showed that ICG/DIST emotional animations (0.11±0.12; mean ± s.e.m.) were recognized significantly worse than those of TKK (0.65±0.08) and ML (0.59±0.05) talking heads. The difference between TKK and ML talking heads failed to reach significance. The ML fearful facial animation was found to be unsuccessful in later analysis (see below), which could have produced unrealistically low mean results for the ML head. Contrast analysis showed that the difference between TKK and ML heads remained non-significant when the fearful facial expressions were ignored (0.66±0.10 *vs.* 0.77±0.06; $F_{1,11}$=1.69, p=0.22). However, the upper 95% confidence limit for the mean recognition score difference between ML and TKK heads (0.11±0.09) was rather high (0.31[1]). Emotions were recognized significantly better from the real actor KH (0.94±0.03) than from any of the talking heads.

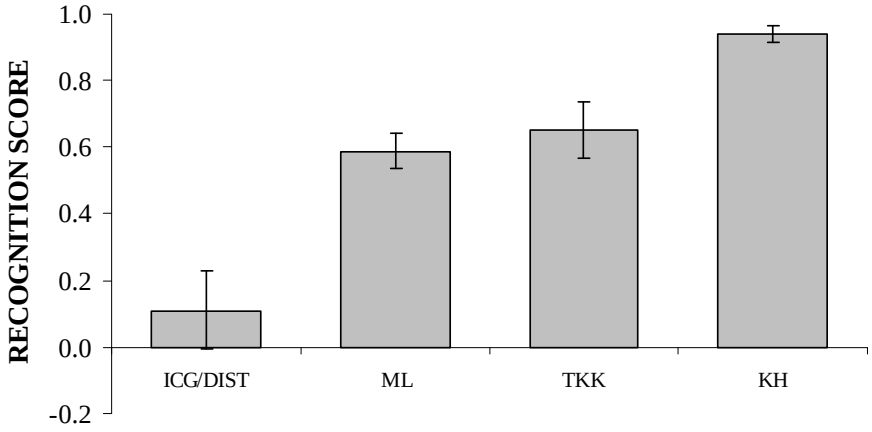

Figure 10 Mean recognition scores (± sem) for different stimulus sets. Chance level corresponds to horizontal axis. Ambiguous recognition level (score 0.6) is marked with dotted line. Please refer to text for the description of significant differences.

[1] Calculated as $\bar{x} + t_{0.05}(11) * s = 0.11 + 2.20 * 0.09 = 0.31$.

Naturalness

Mean *naturalness rates* for facial stimulus sets are shown in Figure 11. A repeated-measures ANOVA revealed significant differences between the stimulus sets ($F_{3,33} = 26.95$, $p<0.0001$). Post-hoc comparison with Newman-Keuls test showed that the ICG/DIST stimuli (0.13±0.05) were evaluated natural less often than those of the other sets, and that the TKK stimuli (0.33±0.09) were evaluated natural less often than those of the ML talking head (0.78±0.04) and the actor KH (0.82±0.07). Naturalness rates between the last two didn't differ significantly.

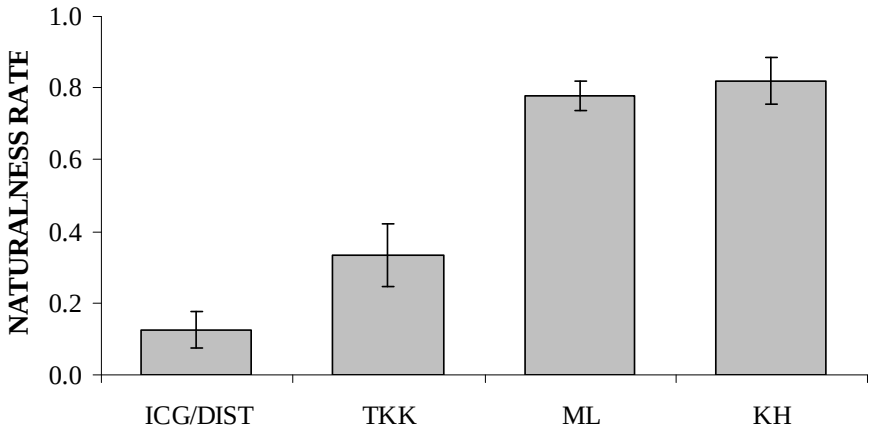

Figure 11 Mean *naturalness rates* (± sem) for different stimulus sets. Please refer to text for the description of significant differences.

Confusions

Mean recognition scores (over individual facial animations) of stimulus sets (Figure 10) were compared to chance level (recognition score 0) with multiple one-tailed t-tests. As a result, all stimulus sets except ICG/DIST exceeded chance significantly ($\alpha^c=0.05/4$). The upper 95% confidence limit for ICG/DIST was 0.37[1]. Stimulus sets other than ICG/DIST were compared to ambiguous recognition level (score 0.6). The only stimulus set exceeding this level significantly ($\alpha^c=0.05/3$) was that of the human actor KH.

For exploring the recognition of emotions from individual facial expressions and animations, their recognition scores were compared to chance level (recognition score 0) with multiple one-tailed t-tests without correction for multiple comparisons. The results

[1] Calculated as $\bar{x} + t_{0.05}(11) * s = 0.11 + 2.20 * 0.12 = 0.37$.

showed that ICG/DIST angry (−0.07±0.25), disgusted (−0.33±0.19) and fearful (−0.63±0.17) facial animations and ML fearful (−0.33±0.21) facial animation failed to exceed chance level significantly. Upper 95% confidence limit was 0.49 for the first of these animations and below 0.14 with each of the remaining animations[1], indicating at best poor recognition falling below the ambiguous recognition level with all of them. Further analysis of *emotion recognition rates* suggested that ICG/DIST angry facial animation produced roughly similar *rates* both for anger and disgust (mean *rates* 0.42±0.16 *vs* 0.25±0.14). ICG/DIST disgusted facial animation was recognized angry more often than disgusted (0.33±0.15 *vs.* 0.08±0.09; t_{11}=1.91, p<0.04), fearful animation was confused similarly with sadness (0.42±0.16 *vs.* 0.00; t_{11}=2.80, p<0.01) and ML fearful facial animation with surprise (0.92±0.09 *vs.* 0.33±0.15; t_{11}=3.02, p<0.006).

Discussion

The overall recognition and naturalness of TKK emotional facial animations were compared to those of two other parametric talking heads and to emotional facial expressions of a skilled human actor. The recognition of emotions failed to reach that of the human actor with any of the talking heads. Naturally, this result can't be generalized directly to all parametric talking heads because only three talking heads were evaluated. Tentatively, the results suggest that more detailed modeling of facial musculature and facial skin than that provided by parametric animation is necessary for producing realistic emotional facial expressions.

Results showed that the overall recognition of emotions exceeded chance level with ML and TKK talking heads but not with ICG/DIST head. Angry, disgusted and fearful ICG/DIST animations were found to be quite unsuccessful. In an ideal case, emotions should be recognized as well from motion-driven animations as from the natural expressions on the basis of them. However, worse performance of motion-driven animations was noted already in the original ICG/DIST evaluation study [159]. This was suggested partly to be due to the fact that it wasn't possible to track upper and lower eyelid movements, which may be important for some emotional facial expressions.

[1] Calculated as $\bar{x} + t_{0.05}(11) * s = -0.067 + 2.201 * 0.253 \approx 0.49$ and $\bar{x} + t_{0.05}(11) * s = -0.33 + 2.20 * 0.21 < 0.14$.

However, also other emotionally salient areas were omitted, such as skin area below lower eyelids, cheeks and skin adjacent to nose[1] [24, 31, 33]. In the motion-tracking, 27 markers were placed on mouth and eye brows. Successful tracking of emotional facial expressions would require more markers on additional locations than those used with ICG/DIST facial animations. Analysis of individual facial animations indicated that ML fearful animation wasn't recognized successfully. The fact that ML fearful facial animation was actually perceived surprised instead of fearful underlines the importance of evaluating facial animations of talking heads in an independent evaluation study.

Comparison of TKK talking head to the other talking heads was of main interest in this study. As expected, TKK facial animations were recognized better and evaluated more natural than ICG/DIST animations. The ML talking head was selected to this study as representing a parametric talking head with realistic and carefully designed emotional facial animations. Respectively, comparison between ML and TKK heads was of special interest. Contrary to expectation, no significant differences were observed, suggesting that TKK emotional facial animations were recognized reasonably well. However, better performance of ML talking head couldn't be ruled out certainly when results from its unsuccessful fearful animation were ignored.

Although the overall recognition results didn't differ significantly between TKK and ML talking heads, the latter was evaluated more natural. This difference may have been mainly due to the used modeling approach: a bottom-up procedure based on predefined FACS prototypes was used with TKK animations whereas the ML animations were modeled top-down based on existing photographs and with an emphasis on esthetical appearance. Plausibly, the ML facial animations were evaluated more natural also because they contained rigid head movements and slightly more complex movement dynamics. This suggests that rigid head movements and rather simple facial expression dynamics can be used to improve the perceived naturalness of emotional facial animations.

In conclusion, the results of this study indicate clearly that TKK animations were in general recognized above chance and better than those of ICG/DIST talking head. The

[1] Note that the MPEG-4 standard [160] itself doesn't include facial animation parameters for the movement of skin adjacent to nose ("nose wrinkling").

results also gave tentative evidence on that TKK animations were recognized as well as those of ML talking head. Closer inspection of results suggested that all individual TKK animations were recognized above chance level. However, because significance level correction for multiple comparisons wasn't used in these comparisons, some animations could have exceeded chance level due to inflated type-1 error rate. Respectively, to confirm the results of individual TKK animations, an additional evaluation with a larger subject sample is necessary. This is provided in study III (Chapter 4.1).

4 ROLE OF MOTION IN RECOGNIZING BASIC EXPRESSIONS

Several studies have found that motion facilitates the recognition of *identity* from facial images, but only when they have been degraded, e.g., by blurring, inverting, pixelating or showing them as negatives [87, 89, 91, 161]. As suggested earlier (Chapter 1.4), it is reasonable, although not trivial, to expect that the same would apply also to the recognition of emotions. This is the main hypothesis studied in the two experiments presented here. Wehrle *et al* [6], who used synthetic three dimensional facial expressions as stimuli, suggested that dynamics improves the recognition of emotions from facial expressions. However, because natural facial stimuli were not used as controls in this experiment, it is not clear whether the effects were specific only for synthetic stimuli. Motion effect might be obtained using typical synthesized faces lacking fine spatial details, but not with non-degraded natural facial stimuli. To confirm this, recognition of static and dynamic natural and synthetic stimuli was studied in the same experiment (Chapter 4.1). The main hypothesis was studied in the second experiment (Chapter 4.2) with static and dynamic natural faces, blurred to different extent.

4.1 Motion and animated basic expressions (study III)

The effect of motion on the recognition and evaluated naturalness of emotions from synthetic and posed natural stimuli was studied. Synthetic stimuli were facial animations generated with TKK talking head (Chapter 2.3.2) and natural stimuli were posed facial expressions selected from CK collection [39] and recorded specifically for this study[1]. Facial expressions selected from EF [5] collection were used as controls. The hypothesis was that the effect of dynamics would occur only with synthetic stimuli, because of the low intensity and the lack of fine spatial details on the talking head. It was expected that dynamics would facilitate the recognition of emotions, but not increase their naturalness.

Recognition of emotions from individual TKK facial animations, with and without additional facial texture, was studied to complement results from the preceding

[1] Note that the TKK collection (Chapter 2.3.1) wasn't used because this study was conducted before the TKK collection was recorded.

evaluation study (Chapter 3.3). In the preceding evaluation, overall recognition level was found to exceed chance level and to equal that of a good-quality talking head [156]; however, individual facial animations weren't studied in detail. The hypothesis of the present study was that the recognition of emotions exceeds chance level with all facial animations. The worse recognition results of all talking heads in comparison to human actor observed in the preceding study suggests pronounced confusions between basic expressions in synthetic emotional facial expressions. It is expected that such confusions would be especially evident with the fearful and disgusted TKK animations because of the typical confusions between disgust and anger on the one hand and fear and surprise on the other (Chapter 1.1). In the preceding study, the effect of facial texture couldn't be evaluated because of other differences between the evaluated talking heads. In the present study, it was expected that additional facial texture would increase the perceived naturalness of facial animations, but not affect the recognition of emotions.

Methods

Research methods follow those described in Chapter 2 with the changes and additions defined here.

Subjects

Subjects were 55 university students (37 males, 18 females; 20-29 years old) from the Helsinki University of Technology (TKK) who participated in the experiment as a part of their studies. All subjects had either normal or corrected vision. All subjects were native speakers of Finnish.

The subjects were asked to fill TAS-20F self-report questionnaire [127] measuring alexithymic personality trait [128], defined as having difficulties in expressing and experiencing emotions. The TAS-20F scores or sub-scores of subjects did not differ statistically from the reference values of Finnish population [129].

Stimuli

In total, the stimuli contained 8 static and 6 dynamic sets of six basic expressions: two static/dynamic sets of synthetic stimuli, four static/dynamic sets of natural stimuli and two static sets of control stimuli. Dynamic sets contained short (mean 0.8 s, range

0.4-1.1 s) video sequences (25 frames per second) showing a transition from a neutral to emotional expression. Static sets contained either original picture material or – if the originals were video sequences – pictures created by selecting the last frame from corresponding video sequences. The synthetic facial expressions were facial animations selected from TKK talking head (TH) (Chapter 2.3.2). The natural facial expressions were posed by human actors and either selected from existing collections or recorded specifically for this study. The TH basic expression prototypes (Appendix B) were used as references for selecting and posing the natural stimuli.

Two static/dynamic sets were selected from the Cohn-Kanade (CK) collection [39] (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 intended prototypes as closely as possible. Since it was not possible to find a suitable full set of all basic emotions from any single actor, stimuli were selected from various actors. Consequently, the two sets contained stimuli from a total of five different actors. Two static/dynamic sets were recorded in TKK, both sets containing stimuli from one actor. The actors (initials JK and VK) were certified FACS coders [162] trained in controlling facial muscles associated with FACS Action Units. The posed facial expressions were based on the TH facial expression prototypes. Because the CK and TKK stimuli hadn't been evaluated by naïve observers before, stimuli from Ekman-Friesen (EF) collection [5] served as control stimuli. Two static sets were selected, 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 [5]. Two static/dynamic sets were selected from TH: textured and non-textured (Figure 6). These two versions were identical except that in the textured version a photograph of a real face (the TKK actor JK) was mapped on the surface of the talking head (cf. Figure 5). The stimulus size was either 22 cm × 17 cm (CK and TKK sets) or 14 cm × 20 cm (EF and TH sets)

Procedure

The subjects were distributed randomly into two groups, which were shown either static or dynamic stimuli sets. As an exception, the static EF sets served as control stimuli for both groups. In addition to other emotional expressions, the static group evaluated also neutral faces showing no emotions[1]. To equalize sample sizes, one subject's data were removed from the first group. The subject with the most deviant mean recognition score was selected. Consequently, both groups contained 27 subjects (18 male, 9 female, mean age 23 years). Group 1 saw 8 static sets of 7 facial expressions (six emotional expressions and neutral face) whereas group 2 saw 6 dynamic sets and 2 static sets of 6 facial expressions (emotional expressions). The TAS-20 scores or sub-scores did not differ significantly between the two groups.

Error correction

Error correction was conducted separately for the two subject groups. As a result, a maximum of one emotional rating was changed in the data of individual subjects in the group 1 (the mean number of changes was 0.2 ratings per subject) and a maximum of two ratings in the group 2 (mean 0.2). No naturalness ratings were modified in either group.

Results

The effect of dynamics

A mixed-design ANOVA was used in evaluating the significance of factors dynamics (static, dynamic), source (posed, synthetic) and expression (six basic expressions) in the recognition of emotional facial expressions and their naturalness evaluations. The CK and TKK sets were pooled together as the posed source (EF was not included as it contained only static stimuli), and the textured and non-textured facial animations were pooled together as the synthetic source. Analysis of naturalness rates showed no significant results for dynamics main effect or its interactions, suggesting that dynamics had no influence on the perceived naturalness. Figure 12 shows the mean recognition scores for static and dynamic posed and synthetic facial expressions. Dynamics clearly improved the recognition of synthetic, but not natural faces. A significant dynamics × source

[1] Note that results for neutral faces aren't described here. Description of these results can be found in [163].

interaction ($F_{1,52}$=32.45, p<0.0001) supported this observation. Analysis of simple effects confirmed that the interaction was due to the significantly (α^c=0.05/2) better recognition of dynamic rather than static synthetic faces (mean ± s.e.m. 0.67±0.03 *vs.* 0.41±0.04; $F_{1,52}$=25.90, p<0.0001), whereas the difference between the recognition of dynamic and static posed stimuli was not significant (0.82±0.02 *vs.* 0.83±0.02). The synthetic facial expressions were recognized significantly worse than posed ones both from static ($F_{1,52}$=160.61, p<0.0001) and dynamic ($F_{1,52}$=21.328, p<0.0001) stimuli.

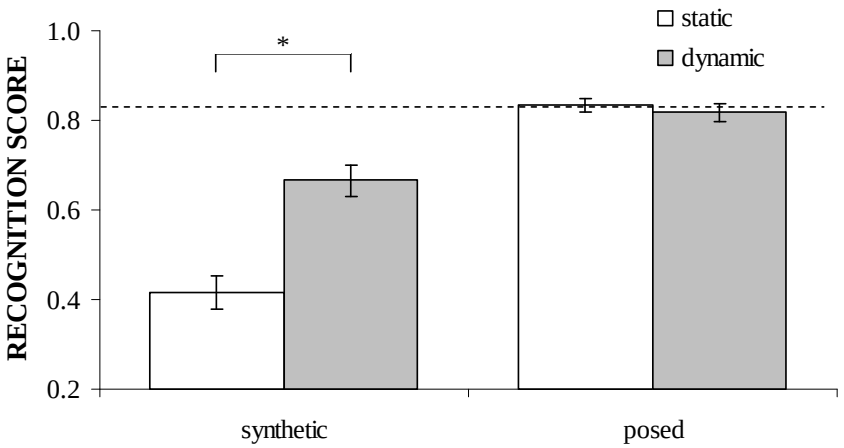

Figure 12 Mean recognition scores (± sem) for static and dynamic synthetic (TH sets) and posed (CK and TKK sets) emotional facial expressions. Asterisk ('*') indicates a significant (p<0.025) difference between static and dynamic scores. The dashed line indicates the mean recognition score (± sem; the shaded region) of the control stimuli (EF sets).

Figure 13 shows recognition scores for static and dynamic versions of the six synthetic facial expressions. Dynamics apparently improved the recognition of angry, disgusted and happy facial expressions. ANOVA revealed a significant dynamics × expression interaction for the synthetic stimuli ($F_{5,260}$=11.58, p<0.0001). Further analysis of simple effects confirmed the significantly (α^c=0.05/6=0.008) better recognition of dynamic rather than static expressions of anger (0.69±0.09 vs. 0.10±0.10; $F_{1,52}$=18.60, p<0.0001) and disgust (0.69±0.08 vs. −0.19±0.13; $F_{1,52}$=32.35, p<0.0001). The difference with happiness failed to reach corrected significance level (0.84±0.05 vs. 0.60±0.09; $F_{1,52}$=5.08, p=0.03). A further analysis of recognition scores with one-tailed t-tests showed that dynamic anger (t_{26}=7.29, p<0.0001) and disgust (t_{26}=9.10, p<0.0001) exceeded chance recognition level (score 0) significantly (α^c=0.05/4) whereas the static versions failed to do so.

Figure 13 Recognition scores (± sem) for static and dynamic *synthetic* expressions. Asterisk ('*') indicates significant difference ($p<0.008$). The dashed line denotes ambiguous recognition level (score 0.6).

Because the stimulus presentation and response times weren't controlled, it is possible although unlikely that the dynamics effect on recognition scores could have been due to the fact that group 2 used more time evaluating the stimuli. To evaluate this hypothesis, the response times were analyzed with a mixed-design ANOVA with factors dynamics, source (posed, synthetic) and expression. There were no significant interactions between dynamics and other factors. A significant dynamics main effect indicated that the response times were significantly longer for dynamic rather than static stimuli (4.4±0.3 s *vs.* 3.5±0.2 s; $F_{1,52}=8.39$, $p<0.006$). Plausibly, response times were longer for dynamic stimuli in comparison to constantly presented static stimuli because subjects preferred to watch at least one full repetition of a video sequence before giving their answer. Importantly, the response latencies were significantly longer for dynamic stimuli both with posed (4.2±0.3 s *vs.* 3.3±0.2 s; $F_{1,52}=9.05$, $p<0.005$) and synthetic (4.6±0.3 s *vs.* 3.7±0.2 s; $F_{1,52}=8.90$, $p<0.005$) facial expressions. This indicates that the better recognition of dynamic over static synthetic stimuli couldn't have been due only to longer response times, because then a similar effect should have been evident also with posed stimuli.

Stimulus source differences

A mixed-design ANOVA with factors group (subject groups 1 and 2) and source (EF, CK, TKK and TH sets) for recognition scores was used to confirm that the overall results weren't influenced by differences between stimulus sources or subject groups. The results are shown in Figure 14. The main effect for source was significant ($F_{3,156}$=80.76, p<0.0001). A post-hoc analysis with Newman-Keuls test showed that TH recognition scores (0.54±0.03) were significantly below those of other sources (mean 0.83±0.01) and that differences between the last weren't significant. Importantly, no significant overall differences were observed between EF control stimuli and CK and TKK stimuli selected for this study.

The group × source interaction was significant ($F_{3,156}$=14.88, p<0.0001), however the recognition scores between groups differed significantly (α^c=0.05/4=0.013) only with TH. This result is equal to the observed dynamics effect with synthetic (TH) stimuli, because the first group saw static and the second dynamic stimuli with the exception of static EF stimuli that were used as controls with both groups. A slight difference between groups was observed also with the control stimuli, which however failed to reach corrected significance level (0.79±0.02 *vs.* 0.86±0.01; $F_{1,52}$=5.39, p=0.02). To evaluate whether this effect could nevertheless have explained the observed difference between static and dynamic TH stimuli, the difference between group 1 and group 2 was compared between EF and TH sets. The result was significant, indicating that the difference between first and second groups was larger with TH than with EF sets ($F_{1,52}$=27.73, p<0.0001). Consequently, any general group-related recognition differences were negligible in relation to the dynamics effect observed with TH.

Naturalness rates for natural stimuli sources were evaluated with a mixed-design ANOVA with factors group (group 1 and 2) and source (EF, CK and TKK). The results showed a significant main effect of source ($F_{2,104}$=59.69, p<0.0001) Post-hoc analysis with Newman-Keuls test indicated that EF sets (0.83±0.02) were evaluated more natural than CK sets (0.59±0.04), which were evaluated more natural than TKK sets (0.50±0.04).

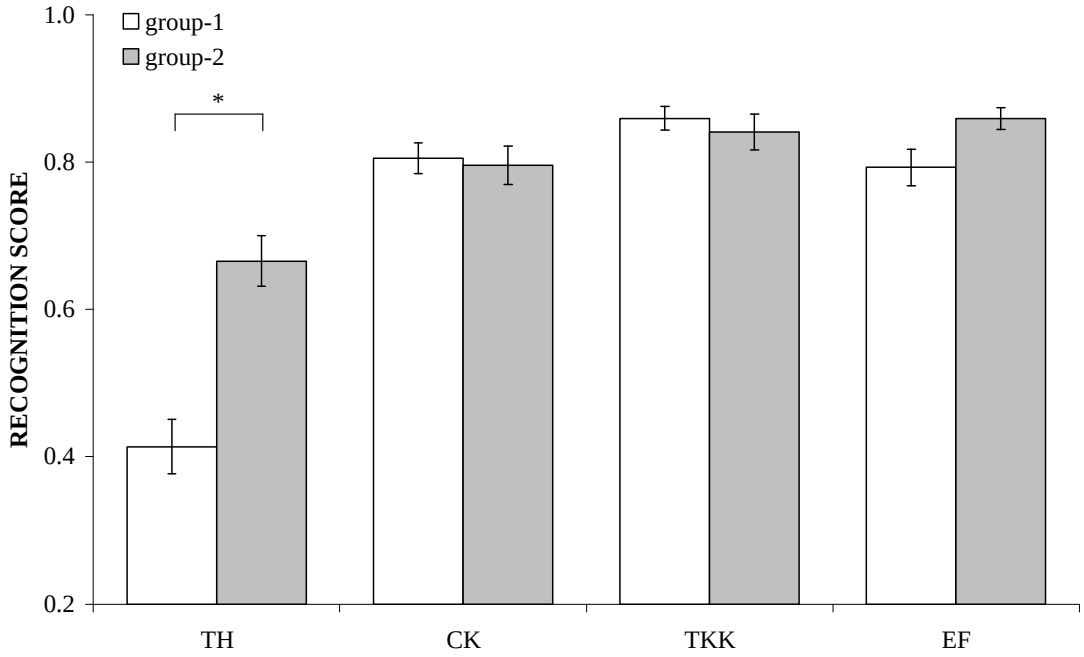

Figure 14 Recognition scores (± sem) for the different subject groups evaluating stimuli from different stimulus sources. Note that group 1 saw only static and group 2 only dynamic stimuli with the exception of EF stimuli that were always static. Asterisk ('*') indicates significant ($p<0.008$) difference.

TKK talking head evaluation

Recognition scores for individual non-textured TH facial animations were compared to chance (score 0) and ambiguous recognition (score 0.6) level thresholds with one-tailed t-tests. Only the non-textured animations were evaluated because the effect of facial texture was studied separately (see below). With all expressions except anger and disgust, results were pooled over static and dynamic stimuli because no significant differences were observed between them. With anger and disgust, only the better recognized dynamic stimuli were considered. The results showed that all emotional facial animations were recognized significantly ($\alpha^c=0.05/6=0.008$) above chance level. On the contrary, the ambiguous level was exceeded significantly only with surprised (0.84 ± 0.05; $t_{53}=5.35$, $p<0.0001$) animation.

For evaluating confusions common to different TKK facial animations, *emotion recognition rates* were compared between target and the most commonly recognized unintended emotions with several two-tailed t-tests. Most common confusions and all secondary confusions with recognition *rates* above 0.10 are shown in Table 5. As in preceding analysis, only the results of dynamic stimuli were considered with angry and

disgusted animations whereas results were pooled over static and dynamic with the remaining animations. Because the analysis of recognition scores had already confirmed that surprise exceeded ambiguous recognition level, multiple comparison correction considered only the remaining animations (α^c=0.05/5=0.01). The results showed that target emotions were recognized significantly more often than most commonly confused emotions with all other facial expressions except fear, with whom surprise received higher recognition *rates* than fear (t_{54}=2.60, p<0.008).

Animated expression	Target	Confusion	Difference	2nd confusion
Fear	0.45±0.07	Surprise: 0.73±0.06	-0.25±0.09	
Sadness	0.65±0.06	Fear: 0.27±0.06	0.38±0.11	Surprise: 0.15±0.05
Disgust	0.67±0.09	Anger: 0.26±0.09	0.41±0.14	
Happiness	0.75±0.06	Surprise: 0.13±0.05	0.62±0.09	Fear: 0.11±0.04
Anger	0.78±0.08	Surprise: 0.11±0.06	0.67±0.13	
Surprise	0.93±0.04	Happin.: 0.20±0.05	0.73±0.08	

Table 5 Mean *emotion recognition rates* (± s.e.m.) for target emotions, most commonly confused non-target emotions and their mean differences for different basic expressions of TH. Last column shows secondary confusions ("2nd confusion") exceeding recognition *rate* 0.10. The results are calculated from the non-textured condition only. All differences are significant (p<0.01).

Textured *vs.* non-textured facial animations

The effect of facial texture on the TH recognition scores and naturalness evaluations were analyzed with a mixed-design ANOVA with factors dynamics, texture (non-textured, textured) and expression (six basic expressions). A significant main effect for texture indicated that the TH was evaluated more natural with facial texture than without it (0.61±0.03 *vs.* 0.53±0.04; $F_{1,52}$=4.57, p<0.04). The texture didn't have significant interactions with other factors.

With recognition scores, a significant interaction between texture and expression ($F_{5,260}$=2.69, p<0.03) showed that the texture effect differed between individual expressions. The further interaction between dynamics, texture and expression was not significant. The recognition scores for textured and non-textured TH facial animations, pooled over static and dynamic stimuli, are shown in Figure 15. Analysis of simple effects showed that facial texture increased the recognition of fearful (0.53±0.08 *vs.* 0.26±0.09; $F_{1,53}$=7.66, p<0.008) facial expression significantly (α=0.05/6=0.008). A slight decrease in textured *vs.* non-textured sad facial expression was also observed, which however didn't reach corrected significance level (0.50±0.08 *vs.* 0.73±0.06;

$F_{1,53}$=6.98, p=0.01). *Emotion recognition rates* for all emotions from textured *vs.* non-textured TH *fearful* expression were studied for evaluating the effect of texture on the recognition of individual emotions further. It was expected that the main changes would occur due to increased recognition of fear, the target emotion, or decreased recognition of surprise, the most often confused emotion. Comparisons with two one-tailed t-tests showed significantly (α^c=0.05/2) decreased recognition of surprise (0.54±0.07 *vs.* 0.72±0.06, t_{52}=2.33, p<0.012) but failed to show significant increase in the recognition of fear (0.56±0.07 *vs.* 0.46±0.07). Analysis of the remaining emotions with two-tailed t-tests showed no significant (α^c=0.05/4) changes.

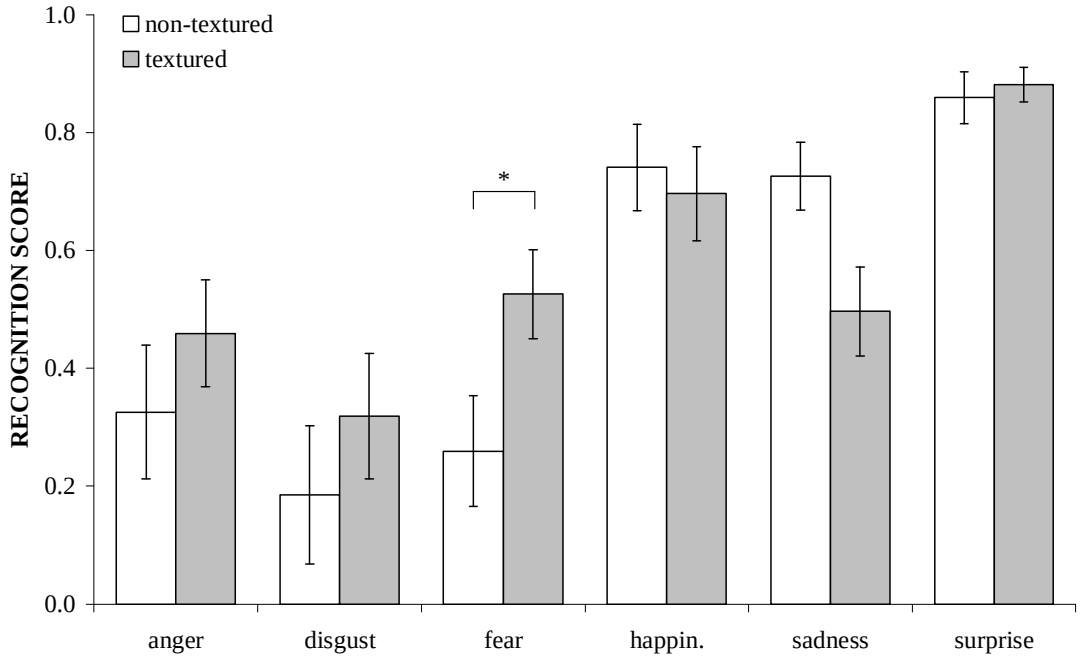

Figure 15 Mean recognition scores (± sem) for the textured and non-textured TH facial animations. Asterisk ('*') indicates significant (p<0.008) differences.

For studying whether texture had any general effects on the recognition of basic emotions, mean *emotion recognition rates* were calculated for all basic emotions so that those facial animations were ignored where a considered emotion was either target or a commonly confused emotion (Table 5). The results, presented in Figure 16, depict the recognition of each basic emotion from facial animations unrepresentative of it. A repeated-measures ANOVA with factors texture (non-textured, textured) and emotion (six evaluated emotions) was used to analyze these general recognition *rates*. The results

showed a significant interaction between texture and emotion ($F_{5,265}$=9.93, p<0.0001). Analysis of simple effects indicated that disgust (0.04±0.01 *vs.* 0; $F_{1,53}$=7.89, p<0.007) and fear (0.23±0.03 *vs.* 0.06±0.02; $F_{1,53}$=26.27, p<0.0001) were recognized significantly (α^c=0.05/6=0.008) more often from textured rather than non-textured animations. Plausibly, the less common recognition of surprise from the fearful animation (see above) was caused by the generally increased perception of fear from the textured talking head.

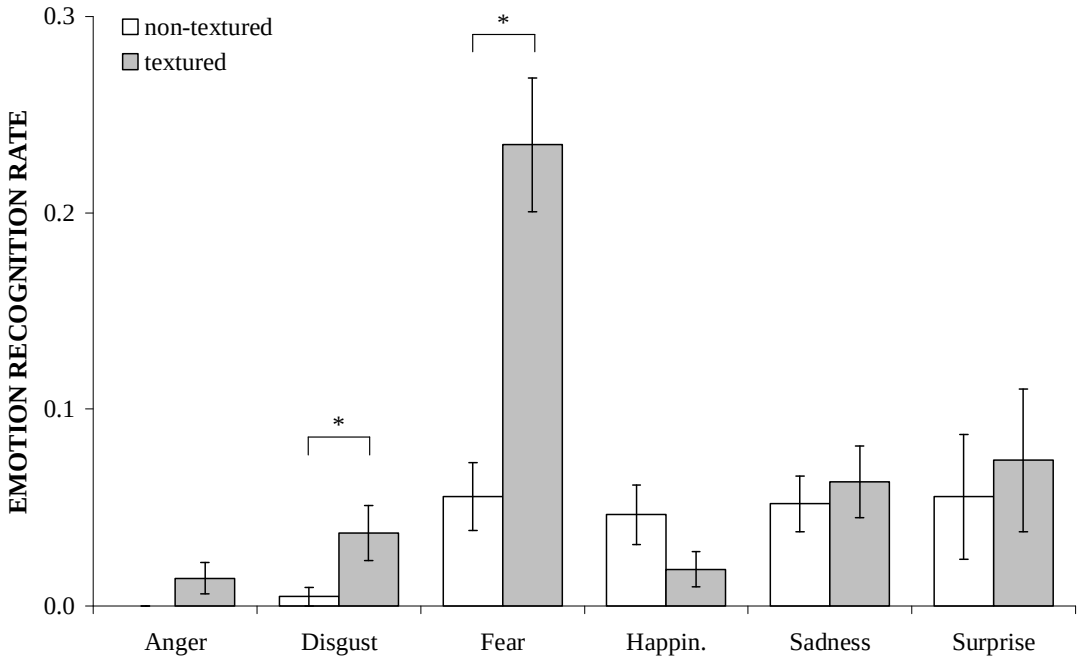

Figure 16 The effect of texture on the general recognition of different basic emotions. The results are shown as mean *emotion recognition rates* (± sem) for different basic emotions, calculated over facial animations unrepresentative of them. Asterisk ('*') denotes significant (p<0.008) differences.

Discussion

Recognition of static and dynamic stimuli was compared both with synthetic facial animations and natural posed facial expressions. A significant difference was observed in favor of the dynamic stimuli in the recognition of angry and disgusted synthetic facial expressions, supporting Wehrle and coworkers' [6] earlier result with synthetic stimuli. The drastic recognition difference between dynamic and static facial expressions is noteworthy, because the dynamics was implemented as a straightforward linear transition from a neutral face to the emotional apex, which was identical with the static expression. No significant difference between static and dynamic stimuli was found with the posed facial expressions. The results are congruent with the initial hypothesis that dynamics facilitates only the recognition of otherwise poorly recognized static stimuli. As already shown in the previous evaluation study (Chapter 3.3), synthetic TKK talking head stimuli were in general recognized worse than posed stimuli. Static angry and disgusted facial animations in particular were recognized at chance level. Similarly, worse recognition of synthetic in comparison to posed stimuli was observed in the previous evaluation with the two other parametric talking heads and in the study by Wehrle *et al* [6]. The worse recognition of synthetic facial expressions in comparison to natural ones was expected to be due to the lack of detailed static features, such as realistic skin wrinkling and bulging. As typical parametric talking heads, the TKK talking head used in this study contained only facial movements and lacked wrinkles, bulges and facial texture changes. Similarly, the synthetic model used by Wehrle *et al* utilized two-dimensional line drawings with a simple facial expression model but without static features.

It is likely that TKK facial animations were recognized worse than posed ones because of their lower intensity in addition to their lack of spatial detail. Because the intensity and spatial accuracy of real faces weren't controlled in this study, it isn't possible to separate their effects on the results. It appears justified to expect that a dynamics effect would be evident also with real faces if their intensities were low or if their spatial accuracy were degraded. The effect of dynamics for extremely subtle facial expressions was observed in a recent study by Ambadar *et al* [96]. In their study, video sequences were selected from the CK collection [39] and clipped to show only 3-6 first video frames from the transition between neutral and emotional faces. The latter claim is studied further in the next

chapter of this thesis and is supported, *e.g.*, by an earlier study where moving points extracted from posed facial expressions were recognized better than similar still point-light presentations [164].

Further evaluation of TKK talking head confirmed that all emotional facial animations were recognized above chance level; however, recognition of anger and disgust exceeded chance only with dynamic facial animations. Ambiguous recognition level was exceeded only by surprised facial animation. Although recognition scores in general failed to exceed the ambiguous recognition level depicting confusion with one unintended emotion, a more detailed analysis of *emotion recognition rates* showed that the intended target emotion was recognized more often than other emotions with all facial animations except fear, which was perceived as surprised. This confusion was similar to but stronger than that observed with well-recognized basic expressions posed by human actors (Chapter 3.2). The apparent conflict between recognition score and *emotion recognition rate* results is explained by the fact that, unlike the latter, the former considers how often the target emotion is in general recognized (cf. Chapter 2.2). On the contrary, comparison between *emotion recognition rates* of target and non-target emotions fails to consider whether the target emotion itself is recognized often enough.

Superimposing facial texture on a talking head increased its perceived naturalness. Unlike expected, the texture also improved the recognition of fearful facial animation by reducing the confusion between fear and surprise. A closer inspection showed that the effect of facial texture was more general in that it increased the recognition of fear also from such facial animations that were not representative of fear. It is suggested here that the increased recognition of fear was due to changed contrast between eyes and facial surface caused by the texture. The same eye model was used both with textured and non-textured talking heads, whereas the facial texture was apparently darker than the original mesh. It is possible that the white sclera of eyes was more pronounced in the talking head with than without facial texture, creating an increased appearance of a fearful facial expression.

4.2 Motion and low-pass filtered posed basic expressions (study IV)

The previous study (Chapter 4.1) showed that dynamics facilitated the recognition of emotions from synthetic but not from real faces. It is obvious that the strength of this conclusion depends on how realistic synthetic stimuli are – the results for real and synthetic stimuli would naturally be indistinguishable from each other if the latter resembled the former closely enough. It was suggested that in this study and the earlier study by Wehrle and co-workers [6], dynamics improved the recognition of only synthetic stimuli because they lacked some details present on real faces. A methodological problem in the previous study was that in addition to the spatial accuracy, also other factors, especially the intensity of facial expressions and the movement dynamics (linear *vs.* natural motion), differed between synthetic and posed stimuli.

In the current study, the role of dynamics in recognizing emotions from degraded stimuli was studied further. To control the extent of degradation accurately, stimuli were blurred by low-pass filtering spatial frequencies (cf. Chapter 0). Earlier studies have evaluated the crucial spatial frequencies for recognizing identity [75-77, 80], audiovisual speech [81] and emotions [82, 84, 86] from faces. Dynamic stimuli have been utilized only in one audiovisual speech recognition study [81]. Importantly, no studies have compared static and dynamic stimuli directly with each other. Furthermore, the recognition of basic emotions from spatially filtered facial expressions hasn't been studied conclusively.

For the current study, low-pass filtering cutoff frequencies (1.8, 3.6, 7.2 and 14.4 c/fw) were selected on the basis of an audiovisual speech recognition study by Munhall *et al* [81], as this was the only previous spatial filtering study with dynamic face stimuli. Earlier low-pass filtering results from an identity recognition task [79] and especially from an emotion recognition task [84] suggest that the recognition of emotions from *static* faces would be degraded at the cutoff frequency 7.2 c/fw. Other similar identity [75, 80, 81] and especially emotion [86] recognition studies suggest that degradation should be evident at cutoff 3.6 c/fw at the latest, which is supported also by earlier band-pass filtering studies highlighting the importance of middle spatial frequencies [76-78,

82]. Making comparisons with the earlier studies is not straightforward, however, because of the various methodological differences between the current and the earlier studies. Most importantly, spatial frequency requirements for emotion, identity and visual speech recognition tasks can be different. Furthermore, large variation in the results could be expected with facial expression research stimuli, as different emotional facial expressions can be expected to rely on different levels of detail. This conclusion was suggested already by the study of Nagayama *et al* [82], where happy faces were recognized better from low spatial frequencies than neutral faces. The hypothesis of the present study was that the general recognition of emotional facial expressions would be degraded at cutoff frequency 3.7 c/fw at the latest, but that this result would differ further between the presented basic expressions.

The main hypothesis of the current study is that no differences between dynamic and static stimuli are evident when the static stimuli are recognized well, but dynamics facilitates the recognition of basic expressions increasingly as the recognition of static stimuli becomes more degraded.

Methods

Research methods follow those described in Chapter 2 with the changes and additions defined here.

Subjects

Subjects were 84 university students (50 males, 34 females; 20-43 years old) from the Helsinki University of Technology (TKK) who participated in the experiment as a part of their studies. All subjects had either normal or corrected vision. All subjects were native speakers of Finnish.

Stimuli

Stimuli contained static and dynamic sets of seven facial expressions (six basic expressions with open- and closed-mouth happiness variants) from four actors (three male and one female; initials KH, NR, SP and TV) selected from TKK collection (Chapter 2.3.1). The actors were selected on the basis of highest overall mean recognition scores in an earlier evaluation (Chapter 3.2). Fearful facial expression of SP was

recognized incorrectly as surprised. However, the SP fear was included to avoid presenting an unequal number of stimuli from different actors, and the results for this stimulus were not included in the main analyses. Dynamic sets contained the original video sequences (mean duration 1.2 s; range 0.7-1.8 s) and static sets contained pictures created by selecting last frames from the sequences.

Blur level	Cutoff frequency (c/face width)	Cutoff frequency (c/deg)	Luminance	RMS contrast
B0	-	-	77	0.74
B1	14.7	3.4	95	0.52
B2	7.3	1.7	94	0.52
B3	3.7	0.8	90	0.52
B4	1.8	0.4	102	0.48

Table 6 Blur levels used in the experiment, their corresponding low-pass filtering cutoff frequencies both on object- (c/face width) and retinal-centered (c/deg) scales and their mean luminance and contrast values.

Original stimuli were converted to a 256 gray-level scale, and resized and cropped to show a constant face width (176 pixels[1]; 61 mm on the screen; 4.4 deg of visual angle at the 80 cm viewing distance) and roughly equally sized light-gray borders around the head. Hair and ears were not masked from the pictures because their impact was considered negligible on the evaluation of emotional facial expressions. Because the borders were kept constant while head shapes, hair styles and other similar factors varied between actors, the resulting picture sizes varied between 265×332 (92 mm × 115 mm) and 288×360 pixels (100 mm × 125 mm). The resized pictures were blurred with a circularly symmetric ideal low-pass filter by convolution method [74], with used luminance values rescaled to 256 grayscale levels. All image manipulations were implemented in Matlab [165]. Four different cut-off frequencies were used (Table 6). Examples of filtering results are presented in Figure 17. The cut-off frequencies were defined primarily on an object scale (cycles per face width) rather than retinal-centered (cycles per degree of visual angle) scale as earlier studies have indicated the former to be more salient for the perception of faces [76, 78, 81].

[1] Measurements were made from the last frame of each video sequence. Face width was measured between the left and right ears' upper attachment points on the head.

Figure 17 Examples of the stimuli (open-mouthed happiness from actor NR). The images are shown from most to least blurred and the original non-blurred one (blur levels B4-B0).

Average luminance (mean values on the 256 gray-level scale) and RMS contrast[1] measures (both corrected for the non-linear luminance response of a CRT monitor[2]) were calculated for each blur condition over all of their dynamic sequences (Table 6). By using ANOVA analyses, significant differences between blur levels were found both with luminance ($F_{4,135}$=13.88, p<0.0001) and contrast ($F_{4,135}$=80.04, p<0.0001). Newman-Keuls post-hoc tests indicated that luminance was significantly lower at blur level B0 than at other levels and higher at B4 than at levels B0 and B3. Contrast was significantly higher at level B0 than at other levels.

Error correction

Original data from 2 subjects were removed because of an unacceptable number (6 and 17) of required error corrections. As a result error correction over the remaining subjects, a maximum of 4 ratings were modified per subject (with a mean of 0.4).

[1] Root mean squares contrast (*RMS*) measures the deviation of luminance values from the mean luminance, calculated with formula (1) where *n* denotes the number of pixels, *l* the luminance of a pixel and l_0 the mean luminance over all pixels (used *e.g.* in [166]).

(1) $$RMS=\sqrt{\frac{\sum\left(\frac{l-l_0}{l_0}\right)^2}{n-1}}$$

[2] Approximated with $l = 255 * \left(\frac{l'}{255}\right)^{\gamma}$ where l' denotes original pixel luminance and $\gamma = 2.2$ on the basis of [167].

Procedure

The subjects were distributed randomly into five groups, which were shown stimuli with different blur levels (Table 7). Stimuli within each group were presented in a randomized order with the constraint that static and dynamic versions of the same facial expression were never presented consecutively. To equalize sample sizes, subjects with the most deviant mean recognition scores within a group were removed so that the size of each group equaled 16 subjects.

Blur-level	Number of subjects			Age	
	Males	Females	Total	Mean	s.e.m.
B0	9	7	16	23.4	0.5
B1	10	6	16	23.3	0.4
B2	9	7	16	24.9	1.3
B3	11	5	16	23.8	0.9
B4	9	7	16	23.8	0.5
	48	32	80		

Table 7 Subject groups and their statistics.

Results

Degradation effect for static stimuli

For the main analysis, original recognition scores were pooled over different actors and expressions (excluding SP fear). The mean recognition scores for static and dynamic stimuli with different blur levels are shown in Figure 18. The recognition of emotions from *static* stimuli blurred at different levels was studied with a between-subjects ANOVA. Post-hoc analysis with Newman-Keuls test, following a significant effect ($F_{4,75}$=86.55, $p<0.0001$), showed that recognition decreased significantly between each subsequent blur level. A significant quadratic trend for all blur levels ($F_{1,75}$=23.00, $p<0.0001$) and a further significant linear trend for levels B0-B3 ($F_{1,75}$=21.40, $p<0.0001$) suggest that the recognition decreased linearly between levels B0-B3 and dropped sharply between levels B3 and B4.

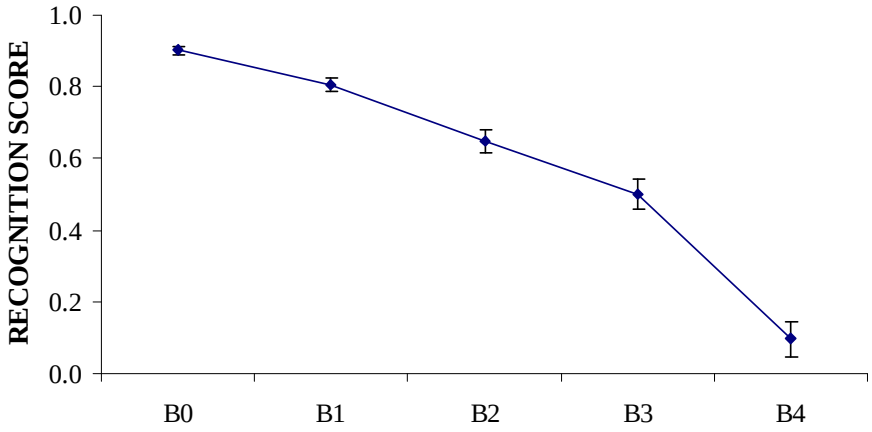

Figure 18 Mean recognition scores (± sem) for *static* stimuli at different blur levels.

A mixed-design ANOVA with factors blur, expression (all expressions) and actor (KH, NR, SP and TV) was used for evaluating whether the effects of blurring differed between different expressions and actors. Results indicated that the interactions blur × expression ($F_{24,450}$=9.47, p<0.0001) and "blur × expression × actor" ($F_{72,1350}$=3.75, p<0.0001) were significant. Contrast tests with one-tailed significance tests were used for comparing the recognition score at each blur level to that of the unblurred level with each basic expression. Similar contrast analyses were repeated separately for each actor. Because the latter analysis was considered exploratory, correction for multiple comparisons was not applied. The results are presented in Table 8. The overall recognition of sadness showed significant (α^c=0.05/(4∗7)=0.002) degradation already at blur level B2, the recognition of anger and disgust at B3, and happiness, fear and surprise at B4. Degraded recognition of closed-mouth happiness at blur levels B2 and B4 but not at their intermediate level B3 is explained by the deviant degradation pattern of two actors (see below).

Expression	Blur level	Static Mean ± s.e.m.	# actors	Dynamic - Static Mean ± s.e.m.	# actors
Anger	B0	0.92 ± 0.04	-	0.03 ± 0.05	-
	B1	0.63 ± 0.16 +	1/4	0.01 ± 0.09	0/4
	B2	0.74 ± 0.05 +	1/4	0.01 ± 0.05	0/4
	B3	0.10 ± 0.09 *	3/4	0.55 ± 0.08 *	3/4
	B4	-0.24 ± 0.09 *	4/4	0.38 ± 0.09 *	2/4
Disgust	B0	0.93 ± 0.02	-	-0.08 ± 0.05	-
	B2	0.74 ± 0.06 +	1/4	0.06 ± 0.05	1/4
	B3	0.09 ± 0.10 *	3/4	0.46 ± 0.12 *	3/4
	B4	-0.35 ± 0.10 *	4/4	0.01 ± 0.08	1/4
Fear	B0	0.63 ± 0.06	-	-0.01 ± 0.03	-
	B4	0.08 ± 0.11 *	3/3	0.43 ± 0.14 *	2/3
Happin. (closed)	B0	0.99 ± 0.01	-	-0.02 ± 0.02	-
	B2	0.35 ± 0.09 *	2/4	0.15 ± 0.10	1/4
	B4	0.30 ± 0.12 *	3/4	0.19 ± 0.13 +	2/4
Happin. (opened)	B0	0.99 ± 0.01	-	-0.01 ± 0.01	-
	B3	0.89 ± 0.06 +	1/4	0.04 ± 0.04	0/4
	B4	0.55 ± 0.07 *	4/4	0.31 ± 0.07 *	4/4
Sadness	B0	0.89 ± 0.03	-	-0.06 ± 0.05	-
	B2	0.36 ± 0.09 *	2/4	0.10 ± 0.07	1/4
	B3	0.24 ± 0.10 *	4/4	-0.11 ± 0.15	0/4
	B4	-0.12 ± 0.12 *	4/4	0.09 ± 0.13	0/4
Surprise	B0	0.91 ± 0.03	-	0.08 ± 0.03	-
	B3	0.79 ± 0.04 +	1/4	0.11 ± 0.05	0/4
	B4	0.57 ± 0.07 *	3/4	0.29 ± 0.08 *	3/4

Table 8 Recognition of static stimuli and difference between dynamic and static stimuli at different blur levels of each basic expression. With static results, recognition scores were compared between other blur levels and level B0 and with difference results, difference scores were compared to zero. Asterisk ('*') denotes overall significance with multiple comparison correction ($p<0.002$). Differences that would have been significant without correction ($p<0.05$) are shown for illustration and are denoted with a plus sign ('+'). Only such blur levels are shown for which a comparison reached significance at either one of these levels. Number of actors ("# actors") refers to individual actors with whom a comparison was significant.

Visual inspection suggested that most facial expressions either followed the general degradation pattern (Figure 18), showed degradation only at the highest blur level or remained roughly constant over all blur levels. For finding obviously deviant degradation patterns, recognition scores were compared between all subsequent blur level increases of individual facial expressions. Null hypothesis was that the recognition would either decrease or remain equal at all comparisons. The alternative hypothesis was that an increase in blur level would, unexpectedly, increase the recognition of emotions. Exploratory data analysis was conducted with one-tailed contrast tests, corrected for the four comparisons conducted for each facial expression ($\alpha^c=0.05/4=0.013$). The results

showed significant increase between blur levels B2 and B3 with closed-mouth happiness of actors NR (−0.60±0.17 *vs.* 0.60±0.19; $F_{1,75}$=94.77; p<0.0001) and SP (0.20±0.23 *vs.* 0.95±0.03; $F_{1,75}$=16.27; p<0.0005). Apparently, with these facial expressions the recognition was degraded at blur level B2 in comparison to both of its neighboring blur levels (Figure 19). Changes in the recognition of individual emotions between blur levels B1 and B2 were studied further by analyzing *emotion recognition rates* for target and the most commonly confused emotions (at level B2) with two-tailed protected t-tests. With actor NR, the results indicated significantly decreased happiness (0.13±0.09 *vs.* 1; t_{75}=13.39, p<0.0001) and increased sadness (0.69±0.12 *vs.* 0; t_{75}=11.39, p<0.0001) recognition and with actor SP, decreased happiness (0.63±0.13 *vs.* 1; t_{75}=4.41, p<0.0001) and increased disgust (0.31±0.12 *vs.* 0; t_{75}=5.18, p<0.0001) recognition.

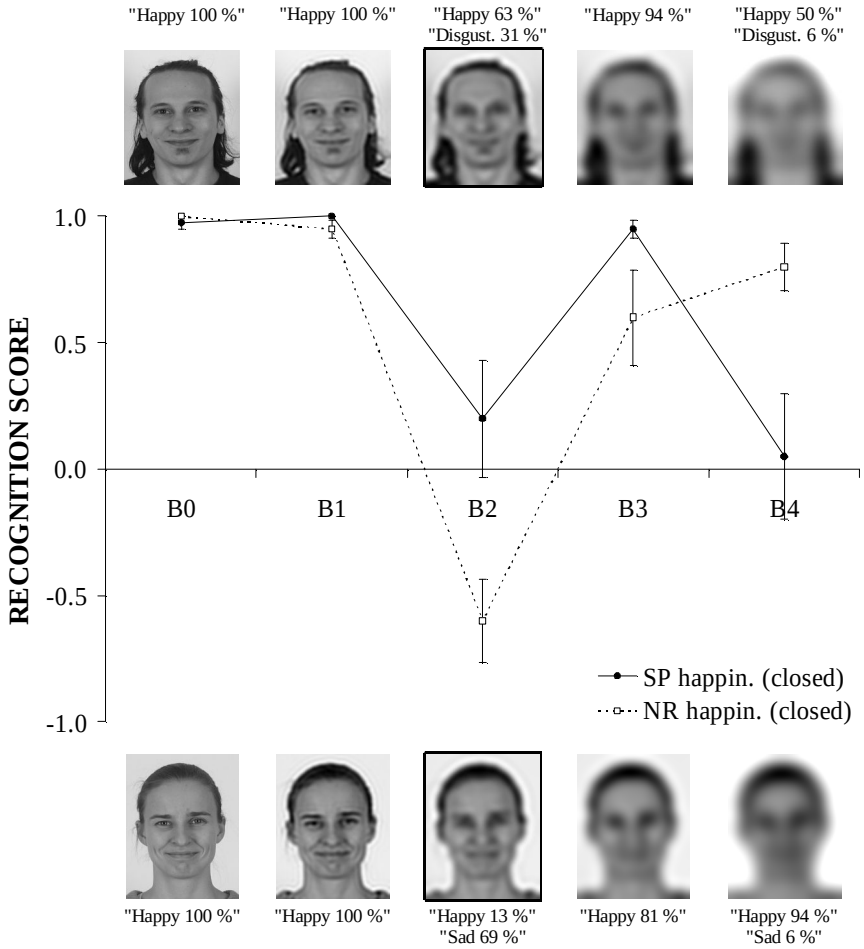

Figure 19 Mean recognition scores (± sem), at different blur levels, for two facial expressions with a deviant degradation pattern. The evaluated stimuli are shown as thumbnail images above (actor SP) and below (actor NR) the figure. Images with unexpected results are shown in frames. Percentages refer to mean *emotion recognition rates* for target and the most commonly confused emotion at blur level B2.

The effect of dynamics

For studying the effect of dynamics, difference scores were calculated where the recognition scores for static stimuli were subtracted from those for dynamic stimuli[1]. The difference scores were analyzed with a between-subjects ANOVA with different blur levels. Figure 20 suggests that although dynamics had no effect with unblurred stimuli, difference between the recognition of dynamic and static stimuli increased constantly as the blur level was increased. This observation was supported by a significant linear trend over blur levels ($F_{1,75}$=52.65, p<0.0001). The main effect of blur level was significant ($F_{4,75}$=13.77, p<0.0001). Planned comparisons showed that the effect of dynamics was significant (α^c=0.05/5=0.01) at blur levels B3 (0.18±0.03, $F_{1,75}$=31.97, p<0.0001) and B4 (0.24±0.04, $F_{1,75}$=60.15, p<0.0001). Post-hoc comparison with Dunnett's test, allowing direct comparison between each blur level and the level B0, showed that the dynamics effect observed at unblurred level B0 (−0.02±0.01) was exceeded significantly only at these levels.

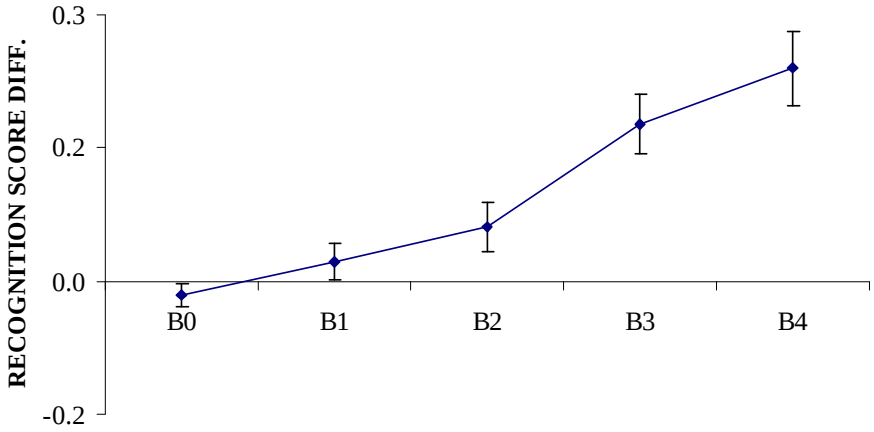

Figure 20 Mean recognition score differences (± sem) between dynamic and static stimuli at different blur levels.

For evaluating whether the dynamics effect differed between expressions and actors, a mixed-design ANOVA was conducted with factors blur, expression and actor. The interactions blur × expression ($F_{24,450}$=3.18, p<0.0001) and "blur × expression × actor" ($F_{72,1350}$=1.54, p<0.004) reached significance, suggesting that the main result varied

[1] Note that because dynamics was a repeated-measures factor, analysis with factor dynamics (static, dynamic) and comparisons between the dynamic and static levels would equal analyses with the used difference score.

between posed basic expressions and further between actors posing them. Contrast tests with one-tailed significance tests were utilized for comparing the difference score at each blur level to that of the unblurred level with each basic expression, and similar procedure was repeated with individual actors. Because the analyses on individual actors were considered exploratory, comparison for multiple comparisons wasn't applied on them. The results are shown in Table 8. Significant (α^c=0.05/(4∗7)=0.002) overall dynamics effect was observed at blur level B3 with anger and disgust and at blur level B4 with fear, opened-mouth happiness and surprise. With closed-mouth happiness and sadness no overall dynamics effect was observed at any blur level. Notably, all significant overall dynamics effects were observed only at blur levels whose recognition was also degraded at the static condition (cf. Table 8).

Visual inspection suggested that with most facial expressions the effect of dynamics either resembled roughly that of the general pattern (Figure 20), peaked before the highest blur level B4 (most angry and disgusted expressions) or remained close to zero at all levels. For finding obviously deviant dynamics effects, recognition score differences were compared to zero at all blur levels of each facial expression with the null hypothesis that the dynamics effect would be zero or positive at all levels. The alternative hypothesis, *i.e.* a negative dynamics effect, would be contrary to the prior expectation. Exploratory data analysis was conducted with one-tailed contrast tests, corrected for the five comparisons conducted for each facial expression (α^c=0.05/5=0.01). The results showed that dynamics decreased recognition significantly with closed-mouth happiness of actor TV at blur level B4 (score difference –0.58±0.24; F=18.06, p<0.0001) (Figure 21). The effects of dynamics on the recognition of individual emotions at blur level B4 was explored by comparing *emotion recognition rates* between static and dynamic stimuli with two-tailed protected t-tests. Dynamics was found to decrease the recognition of happiness (*recognition rates* 0.38±0.13 *vs.* 0.69±0.12; t_{75}=3.36, p<0.002) and to increase the recognition of anger (0.25±0.11 *vs.* 0.06±0.06; t_{75}=3.17, p<0.003) significantly (α^c=0.05/6=0.008).

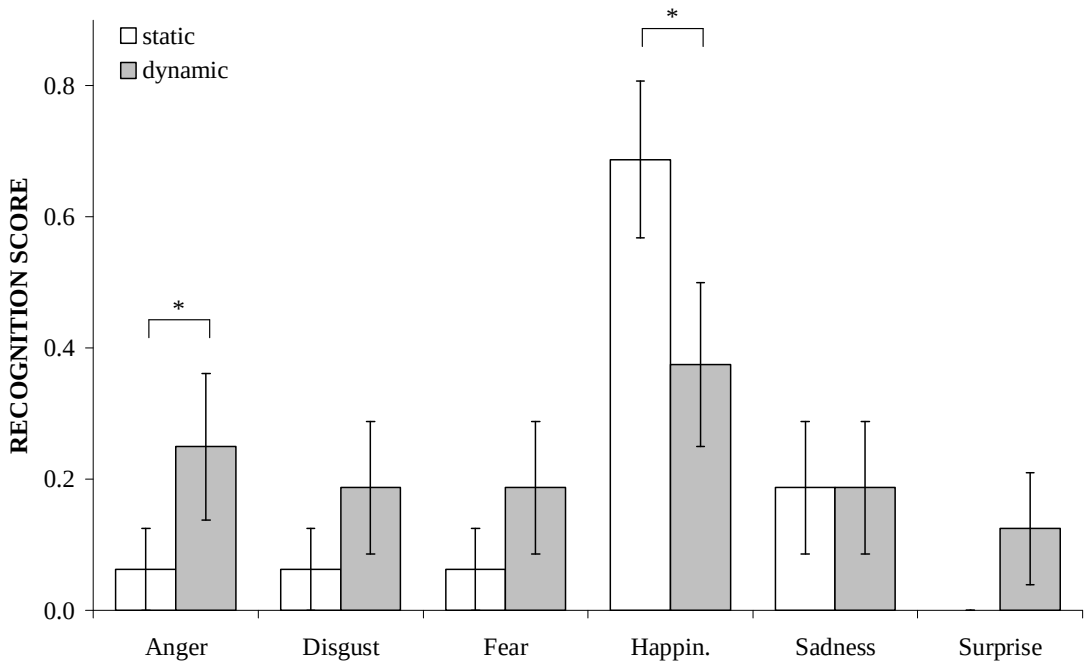

Figure 21 Mean *emotion recognition rates* (± sem) for static and dynamic TV closed-mouth happiness expressions at blur level B3. Significant ($p<0.008$) differences are denoted with an asterisk ('*').

Relation between dynamics and degradation

For evaluating the relationship between dynamics and degradation effects, measures of these effects were recalculated for blur levels B1-B4 of all facial expressions. *Degradation effect* was calculated by subtracting the mean recognition score for a blur level from that of level B0, and *dynamics effect* by subtracting the mean recognition score difference between dynamic and static stimuli at unblurred level B0 from that of a blur level. Pearson's correlation test for the results showed a significant correlation between these measures ($r=0.55$; $t_{110}=6.98$[1], $p<0.0001$), indicating that dynamics facilitated the recognition of emotional facial expressions the more the static stimuli were degraded by blurring. A further comparison between dynamics and degradation effect measures at each specific blur level indicated significant correlations ($\alpha^c=0.05/4$) only at blur levels B2 ($r=0.61$; $t_{26}=3.95$, $p<0.0006$) and B3 ($r=0.69$; $t_{26}=4.83$, $p<0.0001$). The lack of significant correlation between dynamics and degradation effects at blur levels B1 and B4 can be explained by the weak degradation effect at the slightest actual blur level B1 and by a poor recognition of both static and dynamic stimuli at the most severe blur level B4.

[1] Note that this test was calculated over 4 blur levels of all 28 facial expressions, producing a total of 112 data points.

Discussion

Recognition of emotions was studied from static and moving faces blurred by low-pass filtering to different degrees. The main hypothesis was that emotions would be recognized better from dynamic facial expressions in comparison to static ones but only when the static stimuli were degraded enough. This hypothesis was confirmed. The better recognition of dynamic over static stimuli was found to increase linearly when the blur level was increased, reaching statistical significance at the two lowest low-pass filtering cutoff frequencies 1.8 and 3.6 c/fw. Furthermore, the effectiveness of dynamics was found to be correlated with the extent of degraded recognition caused by blurring.

No comprehensive studies on the recognition of basic emotions from *static* low-pass filtered faces have been conducted earlier. In the present study, significant degradation was observed already at the slightest blur level used, with spatial frequency cutoff at 14.7 c/fw. Higher blur level was required for degraded recognition in the study by Costen *et al* [80] where significantly degraded recognition of identity was observed at cutoff frequency 4.5 c/fw (in comparison to cutoff frequency 22.5 c/fw) and in the study by Munhall *et al* [81] where the recognition of dynamic audiovisual speech was degraded (in comparison to unfiltered stimuli) at cutoff frequency 3.7 c/fw. These differences suggest that recognition of emotions depends on higher spatial frequencies than the recognition of identity or audiovisual speech. However, the observed differences are at least partly due to the recognition measure used in this study. The rating task on six emotional scales and the used recognition scoring can be expected to be more sensitive than simple hit rates used in the matching or naming tasks of the previous studies. This suggestion is supported also by the fact that when response times were analyzed in the previous study by Costen *et al*, significant increases were observed already at cutoff frequency 11 c/fw. The present study utilized only low-pass filtering for manipulating the spatial frequency contents of emotional facial expressions. Other methods such as band-pass filtering (cf. existing facial identity [77-79] and facial emotion [82] studies) and adding narrow-band noise (cf. [76] and [84]) could be used in future studies for more accurate evaluations of the specific spatial frequency bands important for perceiving emotional facial expressions.

As expected, both the degradation and dynamics effects varied among the individual facial expression stimuli. Differences were observed between the posed basic expressions, but additional differences existed also between actors. The latter finding reflects the unavoidable heterogeneity of facial expression stimuli that was evident already in the FACS coding of TKK stimuli (Appendix C). Of the static versions of posed basic expressions, fear, happiness and surprise were the least affected by blurring, most of them showing no degradation at all until the highest blur level (cutoff frequency 1.8 c/fw). In comparison to these basic expressions, the recognition of anger, disgust and sadness was degraded at a lower blur level. It appears that (opened-mouth) happiness and surprise were recognizable from low spatial frequencies because all of their prototypes (cf. Appendix B) contained large characteristic changes on the mouth and eye regions. For example, surprise could have been recognized easily from wide vertical mouth opening, (opened-mouth) happiness from opened mouth with upward-turned lip corners and fear from horizontally stretched mouth. In comparison, anger, disgust and sadness contained rather local changes on the face, such as lip tightening, nose wrinkling and lip corner lowering that were more apparent at high than low spatial frequencies. A peculiar degradation pattern observed with the closed-mouth happiness expressions of actors NR and SP, where the recognition peaked down at middle cutoff frequency (cutoff frequency 7.3 c/fw) in comparison to both its higher and lower blur levels (Figure 19), complicates this picture further. Marked confusion with sadness was observed with actor NR and with disgust with actor SP. These confusions are apparently related to non-prototypical facial actions (see Appendix B and Appendix C). Chin wrinkling and rising (AU17) was evident with the closed-mouth happiness of actor NR. This extra activity pushed the middle of lower lip upwards, possibly causing a slight appearance of sadness. On the other hand, the closed-mouth happiness of actor SP contained very slight upper lip rising (AU10), which may have caused an appearance of disgust. Apparently, the relative effects of these factors were strongly pronounced at the middle blur level (7.3 c/fw) in comparison to its adjacent levels (cf. Figure 19). The results suggest that the features related to action units AU10 and AU17 were most evident at spatial frequencies between 3.7-7.3 c/fw but that they were overridden by other facial actions when the frequency band 7.3-14.6 c/fw was included in the filtered spatial frequencies.

An alternative explanation for the discussed unexpected results could be related to a technical issue of using nearly ideal instead of Gaussian low-pass filter (cf. [74]). The advantage of ideal over Gaussian filter is that it passes an exact range of spatial frequencies and suppresses all others. On the other hand, the disadvantage of ideal low-pass filter is that it causes "ringing", *i.e.* the replication of some high spatial frequency contours at the spatial domain (cf. Figure 17; especially the second rightmost image) [74]. Such artifacts could have altered the emotional interpretations of actors' NR and SP happiness expressions at the middle blur level.

Consistently with the main hypothesis, the recognition of anger and disgust were both degraded in static displays and enhanced by dynamics at a lower blur level (cutoff frequency 3.6 c/fw) than fear, (opened-mouth) happiness and surprise (1.8 c/fw). The recognition of sadness and closed-mouth happiness showed minor improvements by dynamics, but they didn't reach statistical significance at any blur level. A significant *negative* effect by dynamics on the recognition of emotions was observed only with closed-mouth happiness of actor TV at the highest blur level (1.8 c/fw) (Figure 21). Visual inspection suggests that a strong vertical larynx movement (not describable by FACS) was characteristic for this expression. This movement was apparently emphasized at the highest blur level due to the lack of other clear features and created a negative appearance.

The main result of this study was that dynamics improves the recognition of basic emotions from low-pass filtered (blurred) but not from unfiltered facial expressions. The overall effect of dynamics was found to increase linearly as the range of low spatial frequencies was narrowed (*i.e.*, the blur level was increased). The specific spatial frequency cutoffs necessary for the recognition to be degraded and the dynamics to improve the recognition were found to depend on the posed basic expression. Degradation and dynamics effects were observed at a slighter blur level with anger and disgust in comparison to fear, opened-mouth happiness and surprise.

5 ASPERGER SYNDROME (AS) AND RECOGNITION OF BASIC EXPRESSIONS

5.1 AS and moving and low-pass filtered posed basic expressions (study V)

Study IV (Chapter 4) confirmed that the recognition of basic emotions from facial expressions is degraded when the stimuli are blurred with low-pass filtering. Facial expression dynamics compensated for this degradation effect. In the current study, the recognition of basic emotions from dynamic *vs.* static degraded (low-pass filtered) facial expression stimuli was compared between adult persons with Asperger syndrome (AS) and matched neurotypical controls. The persons with AS were divided further into two groups on the basis of whether they were diagnosed with prosopagnosia.

The following hypotheses were made for this study. In comparison to control subjects, subjects with AS would have equal recognition accuracies and response times for non-blurred stimuli and lower recognition accuracies for blurred stimuli. No differences were expected between prosopagnosic and non-prosopagnosic subjects with AS in recognition accuracies or response times.

Equal accuracy and response latencies for recognizing basic emotions from non-degraded stimuli were expected between the individuals with AS and neurotypical individuals because earlier studies with ASD adults have suggested deficits only in recognizing mental states more complex than basic emotions (cf. Chapter 1.5). Worse performance in recognizing emotions from low-pass filtered faces is supported by similar result for identity recognition in an earlier spatial frequency study with autistic children by Deruelle *et al* [107]. On the other hand, because this study also suggested improvement during childhood, the performance of adults could equal that of control subjects. Prosopagnosia wasn't expected to have an effect on the recognition of emotional facial expressions, as several patient and brain imaging studies have indicated dissociation between identity and emotion recognition from faces [109]. Furthermore, a study by Hefter and co-workers explicitly comparing emotional facial expression recognition between prosopagnosic and non-prosopagnosic subjects with social

developmental disorders found no relation between facial identity and emotion recognition [101].

No prior hypotheses were made on whether dynamics would be as beneficial for subjects with AS as for controls. In an emotion matching study by Gepner *et al* [115], no difference in the performance accuracy with dynamic stimuli was observed between autistic and non-autistic children. On the other hand, worse recognition in persons with ASD could be expected on the basis of studies indicating that autistic children are in general less sensitive to motion coherence [168] and to complex motion [169] than typically developing children.

Methods

Research methods follow those described in Chapter 2 with the changes and additions defined here.

Subjects

Subjects were 20 adult individuals diagnosed with Asperger syndrome, of whom 9 (5 males and 4 females) were prosopagnosic and 11 (8 males and 3 females) non-prosopagnosic, and 20 neurotypical controls matched on the basis of age (±8 years) and sex. Control subjects were recruited from various sources, including Open University of the University of Helsinki and the Finnish Labour Force bureau. All subjects with AS were diagnosed with the same diagnostic procedure either in Helsinki Asperger Center located in medical center Dextra or in Helsinki University Central Hospital (HUS). Diagnoses were made by skilled clinicians. The diagnostic criteria for AS were based on standard ICD-10 [97] and DSM-IV [98] taxonomies. Prosopagnosia diagnosis depended on criteria adapted from NEPSY test battery[1] [171] and on subject's own personal evaluation. The criteria for AS and prosopagnosia diagnoses have been detailed further in [7]. All subjects were prescreened to exclude schizophrenia, obsessive-compulsory disorders, severe depression and learning disabilities. None of the subjects had psychopharmaceutical medication. Subjects with AS were not prescreened for ADHD

[1] NEPSY is originally designed for children. The used prosopagnosia standards were adapted from those intended for 12-year old children. This procedure was selected because currently, no reliable prosopagnosia tests exist for adults (*e.g.* subjects with developmental prosopagnosia often pass a commonly used Benton Facial Recognition Test [170]). Similar procedure has been used both in clinical and research studies (cf. [7]).

because this disorder appears to be extremely common with AS [7: p. 30]. Earlier records indicated that at least two of the subjects with AS were diagnosed also with ADHD. Control subjects were screened further for autistic spectrum disorders and prosopagnosia. Screening was based on existing medical records with subjects with AS and self-report questionnaire with control subjects. Autistic symptoms of control subjects were evaluated further in an interview with a psychologist, where ASSQ [172] questionnaire intended for screening especially AS and HFA symptoms was used. As a result of screening, one control subject was excluded from a larger initial sample.

All subjects were tested with Wechsler Adult Intelligence Scale-Revised (WAIS-R) [173] producing verbal, performance and full scale intelligence quotient scores and the 20-item Toronto Alexithymia Scale translated in Finnish (TAS-20F) test [127] evaluating alexithymic personality trait. All evaluated subjects had a full-scale IQ higher than 85. Neuropsychological test result were compared between AS and control groups (Table 9), and between prosopagnosic and non-prosopagnosic AS groups (Table 10) with two-tailed t-tests. Results indicated significantly (α^c=0.05/8) higher alexithymia overall scores, as well as factorial scores related to difficulties in identifying and describing feelings, for subjects with AS. Also the third alexithymia factor related to externally oriented thinking was close to significant (t_{38}=1.96; p=0.06). These findings aren't unexpected, because alexithymia is a well-known comorbid disorder for AS (Chapter 1.5). No significant differences were found between prosopagnosic and non-prosopagnosic AS groups. The number of males and females in the prosopagnosic and non-prosopagnosic AS groups didn't differ significantly (χ^2=0.64, *n.s.*).

All subjects had either normal or corrected vision. The subjects were native speakers of Finnish. All subjects were paid for their participation. A written consent was required from all participants. This study was approved by Ethics Committee for Pediatrics, Adolescent Medicine and Psychiatry, and conducted in co-operation with HUS.

		Full AS group (n=20)			Controls (n=20)		
		Mean	SD	Range	Mean	SD	Range
Age		31.9	10.0	18-49	31.2	8.5	19-48
VIQ		110	11	90-127	116	8	104-131
PIQ		113	16	82-144	113	14	85-135
FSIQ		112	13	86-137	116	11	96-134
TAS-20	1)	55	12	31-73	36	6	26-46
TAS-20 F1	1)	21	5	10-29	11	3	7-16
TAS-20 F2	1)	16	6	6-25	9	2	5-14
TAS-20 F3		18	5	9-26	15	5	8-24

Table 9 Means, standard deviations (SD) and ranges of age, IQ and alexithymia measures for AS and neurotypical control groups. The IQ scores include verbal (VIQ), performance (PIQ) and full-score (FSIQ) intelligence quotients, and the alexithymia scores include full TAS-20 scores and its three componential factors F1 (difficulty identifying feelings), F2 (difficulty describing feelings) and F3 (externally oriented thinking). Significant differences (with all $p<0.0001$) between the two groups are marked with an asterisk ('*').

	Prosop. AS group (n=9)			Non-prosop. AS group (n=11)		
	Mean	SD	Range	Mean	SD	Range
Age	35.4	10.6	18-49	29.0	8.9	18-47
VIQ	112	7	102-127	108	13	90-125
PIQ	121	17	95-144	107	13	82-123
FSIQ	117	11	99-137	108	13	86-124
TAS-20	57	12	33-73	54	12	31-68
TAS-20 F1	22	6	10-29	21	4	13-27
TAS-20 F2	17	6	7-25	15	6	6-22
TAS-20 F3	18	5	9-24	18	5	10-26

Table 10 Means, standard deviations (SD) and ranges of age, IQ and alexithymia measures for prosopagnosic and non-prosopagnosic AS groups. The used statistics are same as in **Table 9**. Note that two-tailed t-tests indicated no significant differences between the groups ($p\geq0.07$).

Stimuli

Stimuli contained four static and dynamic sets of angry, disgusted, happy (opened-mouth variant) and fearful facial expressions selected from two male (KH, TV) and two female (NR and MR) TKK collection (Chapter 2.3.1) actors. Note that the actor MR wasn't evaluated in the previous study (Chapter 3.2). Two male and two female actors were selected instead of the three actors with the highest mean recognition scores to avoid any sex-related evaluation differences between the studied groups. Dynamic sets contained the original video sequences (mean duration 1.3 s; range 0.8-1.7 s) and static sets contained pictures created from the last frames of the video sequences.

The original stimuli were processed and low-pass filtered with exactly the same procedure as in the previous experiment (Chapter 4). For the current study, two blur levels were selected on the basis of the previous results so that the first blur level would produce slight or no degradation and the second moderate or severe degradation for the recognition of emotions (Table 11). Respectively, lower cutoff frequencies (more severe blur) were used with fear and happiness than with anger and disgust. The used cutoff frequencies weren't selected at the level of individual actors because the stimuli from MR hadn't been evaluated at different blur levels.

Blur level	Cutoff freq. (c/face width)	
	Anger, disgust	Fear, happin.
Zero	-	
Slight	7.3	3.7
Severe	3.7	1.8

Table 11 Blur levels used in the experiment and their corresponding low-pass filtering cutoff frequencies.

Error correction

Error correction was carried out separately for the AS and control groups. The average number of error corrections per subject was 0.2 for both groups. No more than 2 error corrections per subject were made in either group.

Procedure

All stimuli were evaluated by all subjects. The stimuli were presented in three separate blocks with rest breaks in between, with each block containing stimuli degraded at certain blur level. The blocks were always presented in order B2, B1 and B0, *i.e.* so that all subjects evaluated the severely blurred stimuli in first and the unblurred original stimuli in last block. This fixed order was used because learning effects would have been most detrimental when more blurred stimuli were evaluated **after** their more recognizable versions had already been observed. Some learning effect was to be expected for the slightly blurred and original stimuli, however the blur levels were selected so that their recognition results would already be close to optimal values. The stimuli within each block were presented in random order with the constraint that static and dynamic versions of the same facial expression were never presented consecutively.

Earlier studies (as reviewed in [102, 104, 106]) suggest that subjects with ASD tend to pay attention to abnormal facial features, especially on the lower face. To avoid biasing the subjects' attention on any specific location on faces, the locations of question texts were varied randomly on the screen within a window of 100×100 pixels (35 mm × 35 mm on the screen), centered on the middle of presented stimulus pictures.

The training session contained stimuli not evaluated in the actual experiment (disgusted, fearful and happy facial expressions from actor SP), presented in three similar blocks as in the real experiment. Because of the length of the experiment, it included two rest breaks of at least three minutes. Subjects were encouraged to have longer breaks if necessary and promised refreshments after completing the experiment.

Results

Differences in recognition scores and response times between subjects with AS and controls were studied with a mixed-design ANOVA with factors group (AS, control), blur (none, slight, severe), dynamics (static, dynamic) and expression (anger, disgust, happiness and fear). The results were pooled over individual actors. With response times, the main effect of group and all of its interactions were non-significant. With recognition scores, the interaction group × blur ($F_{2,76}$=3.45, $p<0.04$) reached significance. All other interactions with group were non-significant.

Mean recognition scores for AS and control groups at different blur levels are depicted in Figure 22. The results suggest that there were no differences between AS and control groups at zero and slight blur levels but that the AS group was more degraded at the severe blur level. This observation was confirmed by a significant contrast between the groups at severe blur level ($F_{1,38}$=3.99, $p<0.03$). Because the prior hypothesis was that blurring would influence the subjects with AS more than control subjects specifically at the severe blur level, the contrast analysis used one-tailed significance test and no correction for multiple comparisons.

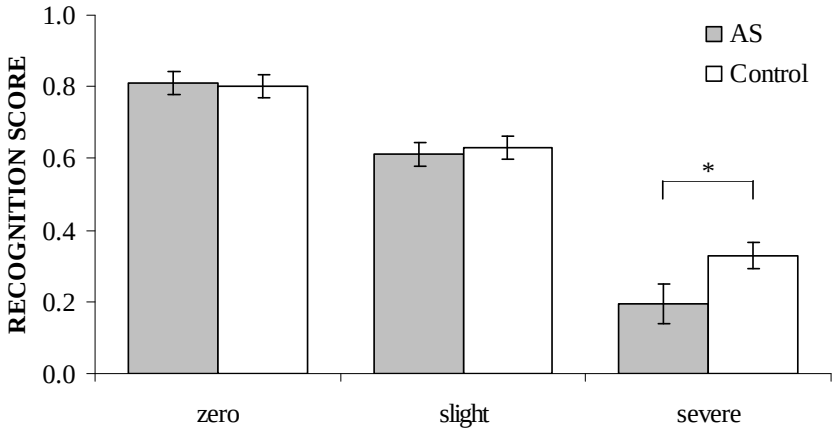

Figure 22 Mean recognition scores (± sem) for AS and control groups at different blur levels, pooled over other factors. Asterisk ('*') denotes significantly ($p<0.05$; one-tailed) lower result for AS group.

Contrast tests confirmed that the difference between AS and control groups at severe blur level didn't vary significantly between evaluated basic expressions or between static and dynamic stimuli. However, because the effect of dynamics was of specific interest, further analysis for static and dynamic stimuli was conducted at severe blur level (cf. Figure 23). Contrast tests confirmed that dynamic stimuli were recognized better than static ones at severe blur level both with AS (0.33±0.05 *vs.* 0.06±0.07; $F_{1,38}=20.47$, $p<0.0001$) and control groups (0.48±0.04 *vs.* 0.17±0.06; $F_{1,38}=27.27$, $p<0.0001$). On the other hand, significantly lower performance with AS in comparison to control group was observed with dynamic ($F_{1,38}=5.38$, $p<0.013$) stimuli. This effect failed to reach significance with static stimuli; however, because static and dynamic stimuli showed similar trend and their difference wasn't significant, it is plausible that the effect existed also with static stimuli.

For studying the significance of differences in recognition scores and response times between prosopagnosic and non-prosopagnosic AS groups, a new mixed-design ANOVA with factors group (non-/prosopagnosic AS, control), blur, dynamics and expression was conducted, where the results were pooled over individual actors. Note that the control group was included in this analysis for increasing statistical power (*i.e.* degrees of freedom for the error term [52]). However, main effects and interactions were tested with contrast analyses where the control group was ignored. As a result, no significant results were observed with the main effect of group or any of its interactions.

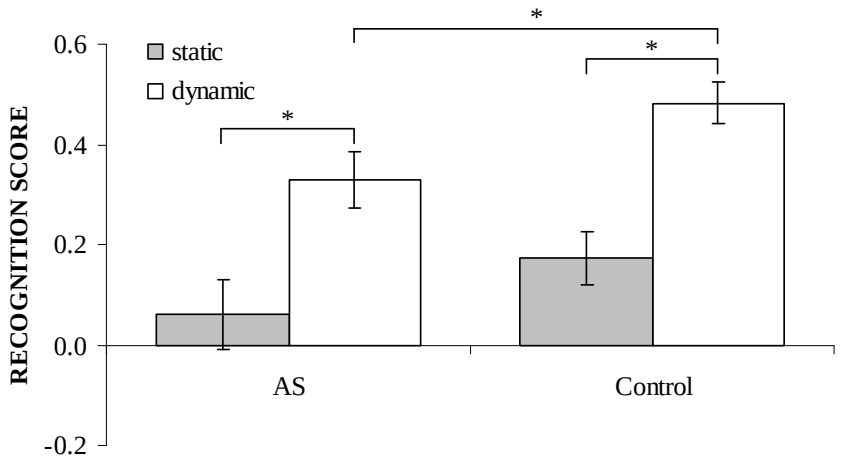

Figure 23 Mean recognition scores (± sem) for AS and control groups and for static and dynamic stimuli at severe blur level, pooled over other factors. Asterisk ('*') denotes significant differences ($p<0.05$; 1-tailed).

Discussion

Results of the current study suggest that subjects with AS recognize at least the studied four basic emotions (anger, disgust, fear and happiness) as well as neurotypical controls from high-fidelity posed basic expressions. This result is congruent with the suggestion that high-functioning adults with ASD are impaired in making complex social and emotional evaluations from the faces of others, but recognize the more simplistic basic expressions typically [118-122]. Concerning the recognition of dynamic *vs.* static basic expressions, the current study suggested no general differences between subjects with and without AS. Subjects with AS recognized dynamic stimuli worse than controls at severe blur level; however, this effect was apparently due to an initial deficit caused by blurring (see below). When the recognition of dynamic and static severely blurred stimuli were compared with each other, subjects with AS were found to benefit as much from observing dynamics as did control subjects. As expected, no significant differences were observed between prosopagnosic and non-prosopagnosic subjects with AS in recognizing emotions from static or moving, degraded or non-degraded facial expressions.

As expected, the results confirmed that adult subjects with AS recognize basic expressions worse from severely blurred stimuli, *i.e.* from low spatial frequencies, than age- and sex-matched neurotypical controls without AS. The result is similar to that of Daruelle *et al* [107] who found that autistic children recognize emotions better from high (above 36 c/fw) than from middle and low spatial frequencies (below 12 c/fw) whereas

an opposite pattern was observed with typically developing children. The results from both of these studies are congruent with the theory of weak central coherence [8] stating that a bias in processing details instead of wholes is characteristic for autistic disorders. Respectively, it is plausible that the worse results in recognizing emotions from blurred facial stimuli were related to a general deficiency in visual information processing rather than a specific deficit in processing faces, facial expressions or emotions.

Why would subjects with AS have difficulties in processing low spatial frequencies? A recent review by Johnson [174] gives a tentative neurological explanation related to subcortical processing of visual (facial) information. Although this review concentrated on face processing, the suggested explanation could apparently be extended also to other types of visual stimuli. As indicated by converging evidence from various neuroimaging studies, subcortical pathway via *superior colliculi* and *pulvinar* to *amygdala* processes low spatial frequencies rapidly and is able to modulate activity in a relatively slower cortical pathway processing high spatial frequencies. Such modulation has been reported for example between amygdala and cortical *fusiform face areas* during processing of faces with direct *vs.* averted gaze direction. Interestingly, several studies have suggested that amygdala functioning is disrupted in ASD already during early childhood. As suggested by Johnson, early disruption in amygdala functioning could lead to weakened processing of low spatial frequencies during development and explain the bias on high *vs.* low spatial frequencies observed both with adolescent and adult persons with ASD. The fact that, in the current study, subjects with AS typically performed worse than neurotypical controls in recognizing emotional facial expressions from low spatial frequencies is compatible with this suggestion. Further evaluation is not possible on the basis of current study that was solely behavioral.

In conclusion, the main result of this study was the confirmation that persons with AS typically recognize emotions worse from blurred facial expressions than neurotypical controls. It was suggested that this effect was due to a general deficit in processing global information (in this study, low spatial frequencies) in comparison to local information (high spatial frequencies). Apparently, the recognition of basic emotions from non-blurred static and from dynamic *vs.* static facial expressions is intact in AS.

6 GENERAL DISCUSSION

The unifying theme of this thesis has been the recognition of basic emotions from dynamic *vs.* static faces. The main hypothesis was that dynamics facilitates the recognition of emotions from facial expressions if, and only if, the recognition of the static versions of the expressions was degraded. This suggestion was motivated by a similar result related to the recognition of identity from faces [87-89, 91, 161, 175]. Earlier studies had shown that emotions are recognized better from dynamic *vs.* static presentations of dots extracted from original faces [87, 94], synthetic facial animations [6], very brief video sequences [96], and from posed unprocessed facial expressions [95]. These studies utilized a wide variety of facial expression stimuli. However, in all but the last of them the static stimuli were in some sense degraded. Therefore, a majority of previous studies is congruent with the presented hypothesis. Results from the one study with deviant results, *i.e.* better recognition of dynamic *vs.* static basic expressions from *non-degraded* stimuli, could reflect peculiarities in the used facial expression stimuli that were recorded specifically for the study and haven't been evaluated in other studies. Most notably, the used dynamic stimuli were long (10 s) in comparison to typical dynamic facial expression stimuli (*e.g.* the length of most dynamic stimuli in [39] being between 1-2 s). It is possible that some confusions could have been more apparent in the static stimuli showing only apexes of facial expressions than in the full movement sequences.

An initial step for studying the main hypothesis was to study the role of motion in recognizing emotions from synthetic stimuli, *i.e.* facial animations produced with a talking head, and from facial expressions posed by human actors (study III). The study by Wehrle *et al* [6] using synthetic stimuli had shown that dynamics facilitates the recognition of emotions, but whether their result could be generalized to naturalistic stimuli was uncertain because posed facial expressions weren't used as control stimuli. Evaluation of the synthetic stimuli indicated that some emotional animations, especially those of anger and disgust, were recognized considerably better from dynamic rather than static presentations. This result confirmed results from the earlier study by Wehrle *et al*. Similar effect wasn't evident with basic expressions posed by human actors. Both in the present and in Wehrle and coworkers' study, facial animations lacked static cues

important for realistic emotional expressions, such as skin wrinkling. Respectively, static versions of emotional animations were difficult to recognize. Consistently with the main hypothesis, observing dynamics compensated for the lack of some potentially important static features in facial animations but didn't affect posed facial expressions that were already easily recognizable.

Evaluation of blurred emotional facial expressions (study IV) confirmed that dynamics is of importance when the recognition of static stimuli is degraded. In the current study, the extent of blurring was quantified by using low-pass filtering with different cutoff frequencies. Higher spatial frequencies are important for perceiving details of visual objects, whereas low frequencies carry information on coarse visual features. To the author's knowledge the present study was the first extensive evaluation of the role of low spatial frequencies in the recognition of basic emotions from facial expressions. Most earlier studies have concentrated on the recognition of identity [75-77, 80] or audiovisual speech [81]. Emotion recognition studies have used only one emotion [82, 86] or have not evaluated the results of different emotions separately [85]. The results showed a linear increase in the facilitating effect of dynamics as the blur level was increased. Further differences were observed between posed basic expressions. Most notably, the facial expressions of fear, happiness and surprise required higher blur level than those of anger, disgust and sadness before any degradation was evident. Such differences were obviously related to the extent of changes on the face, for example fearful, happy and surprised facial expressions containing large characteristic changes on the mouth and eye regions. The differences between basic expressions suggest a fundamental difference between spatial frequencies important for the recognition of identity and emotions from faces. Apparently, identity is always recognized best from a middle spatial frequency band centered approximately at 10 c/fw (cf. Chapter 0). In contrast, the spatial frequencies important for recognizing emotions from faces were found to vary between different basic expressions.

Studies III and IV confirmed the main hypothesis, *i.e.* that dynamics facilitates the recognition of emotions from facial expressions if, and only if, the recognition is degraded with the static versions of the expressions. A modified version of study IV was repeated for studying a new research question related to the evaluation of emotional

facial expression in cognitively high-functioning adults with Asperger syndrome (study V). AS is a developmental neurological disorder belonging to autism spectrum of disorders, characterized by deficits specifically in social communication but without verbal impairments typical for other autistic disorders. Various neurocognitive explanations have been suggested for ASD [8]. Theory of weak central coherence states that an information processing style concentrating on featural instead of configural processing is fundamental for ASD. Respectively, because blurring via low-pass filtering removes featural details producing stimuli with higher demands on configural processing, subjects with AS should be affected more by blurring than neurotypical control subjects. The results of study V confirmed that subjects with AS perform worse than controls with (severely) blurred stimuli. Subjects with AS recognized basic expressions as well as controls from original non-blurred stimuli, indicating that they recognize simple posed expressions typically under normal conditions. No differences between subjects with AS and controls were found in the advantage of evaluating dynamic *vs.* static basic expressions, suggesting that AS involves intact utilization of movement information in recognizing emotions from faces. Prosopagnosia, common for ASD disorders, wasn't found to interact with the recognition of basic expressions.

Synthetic stimuli used in study III were produced by a talking head developed at the Laboratory of Computational Engineering, TKK ("TKK talking head"). Due to shortage of existing dynamic facial expression collections containing posed basic expressions, a new video sequence collection ("TKK collection") was recorded for studies IV and V from six Finnish actors posing FACS action unit combinations. For evaluating these stimuli, the TKK collection was compared to stimuli selected from an existing widely-used picture collection by Ekman and Friesen (EF) [5] (study I) and the TKK talking head was compared to two other parametric talking heads (study II).

Although the actors in TKK collection underwent a long practice period for expressing these combinations, FACS evaluation suggested large differences in the actual facial configurations. This is not unexpected because of the inherent difficulty in posing certain facial expression configurations exactly. Because of the heterogeneity of the stimuli, some differences between individual actors were observed for example in the effect of blurring (study IV). Unfortunately, the heterogeneity of emotional facial expression

stimuli appears inevitable. The use of spontaneous instead of posed emotional facial expressions obviously wouldn't solve this problem as the former are by definition less controlled than the latter.

The evaluation of TKK collection (study I) showed that its basic expressions were recognized as well as those selected from EF collection but were evaluated less natural. Presumably, this difference was due to the careful selection of EF stimuli that was conducted over several years from a large initial sample of pictures. Due to practical limitations, emotional facial expressions from all recorded actors were included into the TKK collection. Extensive selection of stimuli hasn't been utilized in any existing freely available basic expression collection [39, 57, 60, 71, 72]. Careful selection procedure based for example on FACS action units and/or evaluation study with subjects would presumably improve both the distinctiveness of emotional content and evaluated naturalness of available research material.

The facial animations produced with TKK talking head were found to be recognized reasonably well in comparison to those provided by Linköping University [157] and University of Geneva ("MIRALab talking head") [156]. In general, facial animations would appear as an ideal solution for obtaining homogeneous emotional facial expression stimuli. In existing facial expression studies, computer animation has been used for manipulating existing facial expressions of basic emotions, *e.g.*, for creating emotional facial expressions with exaggerated intensity [176, 177], blended emotional facial expressions [45] and artificial movement [55, 178]. With a sophisticated facial animation model, emotional facial expressions could be generated fully automatically with total control over facial configurations, the intensity of facial expressions, head position and various other variables. Evaluation results from the three talking heads were not encouraging, however, as none of them reached the level of a human actor with well-recognized posed basic expressions. Results from only three parametric talking heads must be interpreted with caution; however, it is plausible that realistic facial animation would require more sophisticated modeling of facial skin and its underlying musculature than that available in parametric animation. In general, evaluation by naïve human subjects appears to be important for confirming that facial animations are recognized as intended by the animators. For example, the fearful facial expression of the otherwise

well recognized animations of the MIRALab talking head was judged surprised rather than fearful by most subjects.

All of the presented studies used a rating task where subjects evaluated their agreement on how well each of the basic emotions described a presented expression. The rating task together with the used scoring methods provided detailed information both on how distinctively the intended basic emotions were recognized and on their confusions with other emotions. On the other hand, the used 7-step agreement scale with "uncertain" answer fixed in the middle of the scale had some potential problems. First of all, the used scale offered fewer positive answers (those above the uncertain answer) than an equivalent intensity scale ranging from not felt to strongly felt emotion used in most earlier studies [6, 19, 20, 49, 55, 56]. Consequently, the used scale could have been less sensitive to subtle differences between basic emotions such as the existence of fear *vs.* surprise in fearful faces. This could have lead to the perception of fearful pictures selected from Ekman-Friesen collection [5] as blends of fear and surprise in study II; however, it is unlikely that this would have affected the significant results observed in other studies. Secondly, the possibility for giving an uncertain answer could have affected the results of studies IV and V with blurred stimuli. It is plausible that variation in subjects' performance would have decreased in these studies if the scale would have contained an even number of response options with no option for uncertainty, forcing the subjects to select whether an emotion was or was not present.

In a sense, the recognition scoring used in this thesis and typical recognition accuracy measures used in earlier studies are based on an invalid assumption of straightforward relation between basic expressions and basic emotions. The existence of common confusions between basic expressions is well known, and is apparently related to similar facial expression components between different basic expressions. However, as long as the ambiguity between basic expressions is acknowledged, using recognition scoring based on distinctiveness evaluation can be justified. For example, it makes perfect sense to compare how distinctively basic expressions are recognized from facial expression stimuli selected from different collections.

The present thesis has confirmed that dynamics improves the recognition of basic emotions from (degraded) facial expressions, but hasn't considered the underlying

mechanisms for why this is so. The facilitating effect of motion certainly isn't restricted only to the recognition of faces or facial expressions. For example, classic demonstrations have shown that a walking person can be recognized from moving dots but not from stationary displays and that inanimate objects can be recognized from a few moving but not from stationary dots (see [179] for a review). Respectively, it is rather trivial that the better recognition of dynamic *vs.* static basic expressions is at least to some extent due to general motion perception that isn't specific to faces or facial expressions. On the other hand, it is possible that moving faces would contain also supplemental information related specifically to emotions. Similar proposals have been made for the recognition of identity from faces ([ibid]). Hill and Johnston [88] have shown that gender and identity can be recognized from whole-head and facial movements extracted from human actors and replicated on an animated talking head, suggesting that gender and identity may be characterized by certain head movements and facial expressions. It is conceivable that similarly, certain movements could characterize different emotional states.

Recently, different hypotheses for the facilitating effect of motion in recognizing emotions from very brief facial expression movement sequences have been studied by Ambadar and coworkers [96]. They managed to exclude explanations related to the larger amount of information contained in video sequences rather than pictures, and to the facilitation of configural processing, *i.e.* enhanced processing of relations between individual facial features. They also discarded the existence of emotion-specific dynamic information by showing that when the first and last frames from original video sequences were shown in succession ("first-last presentation"), as large dynamics effect was observed as with the originals. The authors concluded that the facilitating effect of dynamics was due to enhanced change perception provided by the comparison between emotional and neutral faces. This is a viable hypothesis; however, it is questionable whether the results can be generalized from brief to full emotional facial expression movement sequences. In their study, the stimuli consisted of 4-7 frames[1] from the beginning of full emotional facial expression sequences selected from the Cohn-Kanade collection [39]. Note especially that the used dynamic stimuli contained only 2-5

[1] Presented by showing the first frame for 500 ms, the consequent frames for 100-200 ms (3-6 frames at a frame rate of 30 frames/s) and the last frame until a response was received from a subject.

additional frames in comparison to the first-last presentations. It is uncertain whether such brief presentations could replicate possible emotion-characteristic movements contained in a full video sequence. It is suggested here that for studying this issue further, future studies should replicate the research procedure used in study IV of the present thesis with a further first-last presentation condition adopted from the study by Ambadar and co-workers. In this kind of study, null hypothesis would be that successive presentation of only the first and last frames from a video sequence would increase the recognition of degraded stimuli as much as observing the full video sequence. An opposite result would support the existence of emotion-characteristic facial movements.

In conclusion, the present thesis confirmed that dynamics improves the recognition of basic emotions from degraded but not from well-recognized facial expressions of emotions. The degraded stimuli included facial animations that lacked accurate spatial details and posed facial expressions that were blurred by low-pass filtering. Evaluation studies confirmed that the used basic expression stimuli were recognized well in comparison to other existing facial expression stimuli. The low-pass filtering results indicated that, unlike identity recognition depending on a constant range of spatial frequencies, different spatial frequency bands are crucial for different basic expressions. A further study showed that Asperger syndrome involves a deficit in recognizing emotions from low spatial frequencies but no deficit in processing dynamic *vs.* static facial expressions.

REFERENCES

1. Ekman P, *About brows: emotional and conversational signals*, in *Human Ethology*, M von Cranach, K Foppa, W Lepenies, et al. (ed.), 1979, Cambridge University Press: Cambridge. pp 169-249.
2. Argyle M & Cook M, *Gaze as part of the sequence of interaction (chapter 5)*, in *Gaze and Mutual Gaze*, 1976, Cambridge University Press: Cambridge. pp 98-124.
3. Ekman P, Friesen WV, Ellsworth P, *What emotion categories or dimensions can observes judge from facial behavior?*, in *Emotions in the Human Face*, P Ekman (ed.), 1982, Cambridge University Press: London. pp 39-55.
4. Ekman P, *Cross-cultural studies of facial expression*, in *Darwin and Facial Expression*, P Ekman (ed.), 1973, Academic Press Inc.: London. pp 169-222.
5. Ekman P & Friesen W, *Pictures of Facial Affect*. 1978, Consulting Psychologists Press: Palo Alto, CA.
6. Wehrle T, Kaiser S, Schmidt S, Scherer KR, *Studying dynamic models of facial expression of emotion using synthetic animated faces.* Journal of Personality and Social Psychology, 2000, **78**(1): 105-119.
7. Nieminen-von Wendt T, *Asperger Syndrome: Clinical, Neuroimaging and Genetical Findings*. Department of Child Neurology, University of Helsinki.
8. Hill EL & Frith U, *Understanding autism: Insights from mind and brain.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences., 2003, **358**(1430): 281-289.
9. Darwin C, *The Expression of the Emotions in Man and Animals*, ed. P Ekman. 1872/1998, Oxford: Oxford University Press.
10. Solomon RC, *Back to basics: On the very idea of "basic emotions".* Journal for the Theory of Social Behaviour, 2002, **32**(2): 115-158.
11. de Boulogne D, *The Mechanism of Human Facial Expression*. Studies in emotion and social interaction, ed. A Cuthbertson. 1862/1990, Cambridge: Cambridge University Press.
12. Griffiths PE, *Basic Emotions, Complex Emotions, Machiavellian Emotions*, in *Philosophy and the Emotions*, A Hatzimoysis (ed.), 2003, Cambridge University Press: Cambridge.
13. Ortony A & Turner TJ, *What's Basic About Basic Emotions?* Psychological Review, 1990, **97**(3): 315-331.
14. Ekman P, *Expression and the nature of emotion*, in *Approaches to Emotion*, K Scherer and P Ekman (ed.), 1984, Lawrence Erlbaum: Hillsdale, N. J.
15. Ekman P & Friesen WV, *A new pan-cultural facial expression of emotion.* Motivation and Emotion, 1986, **10**(2): 159-168.
16. Ekman P, *Facial Expressions*, in *Handbook of Cognition and Emotion*, T Dalgleish and M Power (ed.), 1999, John Wiley & Sons Ltd.: New York.
17. Russell JA, *Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies.* Psychological Bulletin, 1994, **115**(1): 102-141.

18. Ekman P & Friesen WV, *Constants across cultures in the face and emotion.* Journal of Personality and Social Psychology, 1971, **17**: 124-129.
19. Ekman P, Friesen WV, O'Sullivan M, Chan A, Diacoyanni-Tarlatzis I, et al., *Universals and cultural differences in the judgments of facial expressions of emotion.* Journal of Personality and Social Psychology, 1987, **53**(4): 712-717.
20. Russell JA, *Negative results on a reported facial expression of contempt.* Motivation and Emotion, 1991, **15**(4): 281-291.
21. Haidt J & Keltner D, *Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition.* Cognition and Emotion, 1999, **13**(3): 225-266.
22. Ekman P, *Universals and cultural differences in facial expressions of emotion*, in *Nebraska Symposium on Motivation*, JK Cole (ed.), 1971, University of Nebraska Press: Lincoln.
23. Ekman P, *Basic Emotions*, in *Handbook of Cognition and Emotion*, T Dalgleish and M Power (ed.), 1999, John Wiley & Sons Ltd.: Sussex, U.K.
24. Ekman P & Friesen W, *Unmasking the face. A guide to recognizing emotions from facial expressions*. 1975, Palo Alto, CA: Consulting Psychologists Press.
25. Ekman P, *Facial expression and emotion.* American Psychologist, 1993, **48**(4): 384-392.
26. Shaver P, Schwartz J, Kirson D, O'Connor C, *Emotion knowledge: further exploration of a prototype approach.* Journal of Personality and Social Psychology, 1987, **52**(6): 1061-86.
27. Wagner HL, *Methods for the study of facial behavior*, in *The Psychology of Facial Expression*, JA Russell and JM Fernández-Dols (ed.), 1997, Cambridge University Press/Maison des Sciences de l'Homme: Cambridge/Paris. pp 31-54.
28. Tomkins SS, *Script theory: differential magnification of affects.*, in *Human Emotions: A Reader*, JM Jenkins, K Oatley, and NL Stein (ed.), 1998, Blackwell Publishers: Oxford. pp 209-218.
29. Ekman P, *Should we call it expression or communication?* Innovations in Social Science Research, 1997, **10**(4): 333-344.
30. Ekman P, Levenson RW, Friesen WV, *Autonomic nervous system activity distinguishes between emotions.* Science, 1983, **221**(1208-1210).
31. Ekman P, Friesen W, Hager J, *Facial Action Coding System*. 2nd ed. 2002, Salt Lake City: Research Nexus eBook.
32. Sayette MA, Cohn JF, Wertz JM, Perrott MA, Parrott DJ, *A psychometric evaluation of the Facial Action Coding System for assessing spontaneous expression.* Journal of Nonverbal Behavior, 2001, **25**: 167-186.
33. Ekman P, Friesen W, Hager J, *Facial Action Coding System: Investigator's Guide*. 2nd ed. 2002, Salt Lake City: Research Nexus eBook.
34. Ekman P, Irwin W, Rosenberg EL, *EMFACS-7 (unpublished manuscript)*. 1994.
35. Internet source: *DataFace*. URL: http://face-and-emotion.com/dataface
36. Carroll JM & Russell JA, *Do facial expressions signal specific emotions? Judging emotion from the face in context.* Journal of Personality and Social Psychology, 1996, **70**(2): 205-218.
37. Ekman P, *Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique.* Psychological Bulletin, 1994, **115**(2): 268-87.

38. Tomkins SS & McCarter R, *What and Where Are the Primary Affects? Some Evidence for a Theory.* Perceptual and Motor Skills, 1964, **18**: 119-58.
39. Kanade T, Cohn JF, Tian Y. *Comprehensive database for facial expression analysis.* in *4th IEEE International Conference on Automatic Face and Gesture Recognition*. 2000.
40. Dailey MN, Cottrell GW, Padgett C, Adolphs R, *EMPATH: a neural network that categorizes facial expressions.* J Cogn Neurosci, 2002, **14**(8): 1158-73.
41. Izard CE, *Innate and universal facial expressions: Evidence from developmental and cross-cultural research.* Psychological Bulletin, 1994, **115**(1): 102-141.
42. Russell JA, *Negative results on a reported facial expression of contempt.* Motivation and Emotion, 1990, **15**: 281-291.
43. Rosenberg EL & Ekman P, *Conceptual and methodological issues in the judgment of facial expressions of emotion.* Motivation and Emotion, 1995, **19**: 111-138.
44. Russell JA, *Facial expressions of emotion: What lies beyond minimal universality?* Psychological Bulletin, 1995, **118**(3): 379-391.
45. Young AW, Rowland D, Calder AJ, Etcoff NL, Seth A, et al., *Facial expression megamix: tests of dimensional and category accounts of emotion recognition.* Cognition, 1997, **63**(3): 271-313.
46. Russell JA, *A circumplex model of affect.* Journal of Personality and Social Psychology, 1980, **39**(6): 1161-1178.
47. Woodworth RS & Schlosberg H, *Experimental Psychology: Revised edition*. 1954, New York: Henry Holt.
48. Stanislawski KS, *La formation de l'acteur [Building a Character]*. 1975, Paris: Payot.
49. Gosselin P, Kirouac G, Doré FY, *Components and recognition of facial expressions in the communication of emotion by actors.* Journal of Personality and Social Psychology, 1995, **68**(1): 83-96.
50. Elfenbein HA & Ambady N, *When familiarity breeds accuracy: Cultural exposure and facial emotion recognition.* Journal of Personality and Social Psychology, 2003, **85**(2): 276-290.
51. Wagner HL, *On measuring performance in category judgment studies of nonverbal behavior.* Journal of Nonverbal Behavior, 1993, **17**(1): 3-28.
52. Howell DC, *Fundamental Statistics for the Behavioral Sciences*. 2003: Wadsworth Publishing.
53. Metsämuuronen J, *Tutkimuksen Tekemisen Perusteet Ihmistieteissä (in Finnish)*. 3rd ed. 2005, Jyväskylä: Gummerus Kirjapaino Oy.
54. Bartneck C & Reichenbach J, *Subtle emotional expressions of synthetic characters.* Journal of Human-Computer Studies, 2005, **62**: 179-192.
55. Kamachi M, Bruce V, Mukaida S, Gyoba J, Yoshikawa S, et al., *Dynamic properties influence the perception of facial expressions.* Perception, 2001, **30**: 875-887.
56. Parker JDA, Taylor GJ, Bagby RM, *Alexithymia and the recognition of facial expression of emotion.* Psychotherapy and Psychosomatics, 1993, **59**: 197-202.

57. Lyons MJ, Akamatsu S, Kamachi M, Gyoba J. *Coding facial expressions with gabor wavelets.* in *Third IEEE International Conference on Automatic Face and Gesture Recognition*. 1998. Nara Japan: IEEE Computer Society.
58. Stanislaw H & Todorov N, *Calculation of signal detection theory measures.* Behavior Research Methods, Instruments, & Computers, 1999, **31**(1): 137-149.
59. Green DM & Swets JA, *Signal Detection Theory and Psychophysics*. 1966, New YOrk: J. Wiley and Sons, Inc.
60. Pantic M, Valstar MF, Rademaker R, Maat L. *Web-based database for facial expression analysis.* in *IEEE International Conference on Multimedia and Expo (ICME)*. 2005. Amsterdam.
61. Internet source: *emotion-research.net*. URL: http://emotion-research.net/wiki/Databases
62. Internet source: *Face Recognition Homepage*. URL: http://www.face-rec.org/databases/
63. Internet source: *Google search engine*. URL: http://www.google.com
64. Gross R, *Face Databases*, in *Handbook of Face Recognition*, SZ Li and AK Jain (ed.), 2005, Springer-Verlag: New York.
65. Douglas-Cowie E, Cowie R, Schröder M. *A new emotion database: considerations, sources and scope*. in *ISCA Workshop on Speech and Emotion.* 2000. Northern Ireland.
66. Sim T, Baker S, Bsat M, *The CMU pose, illumination, and expression database.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, **25**(12): 1615-1618.
67. Internet source: *Psychological Image Collection at Stirling (PICS)*. University of Stirling Psychology Department. URL: http://pics.psych.stir.ac.uk/
68. Georghiades AS, Belhumeur PN, Kriegman DJ, *From few to many: Illumination cone models for face recognition under variable lighting and pose.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, **23**(6): 643-660.
69. Martinez AM & Benavente R, *The AR Face Database*. CVC Technical Report #24. June, 1998.
70. O'Toole AJ. *Personal communication*, 15th February 2006.
71. O'Toole AJ, Harms J, Snow SL, Hurst DR, Pappas MR, et al., *A video database of moving faces and people.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, **27**(5): 812-816.
72. Battocchi A, Pianesi F, Goren-Bar D. *A First Evaluation Study of a Database of Kinetic Facial Expressions (DaFEx)*. in *International Conference on Multimodal Interfaces (ICMI)*. 2005. Torento, Italy: ACM Press.
73. Goldstein EB, *Sensation and Perception*. 5th ed. 1999, Pacific Grove, CA: Brooks/Cole Publishing Company.
74. Gonzales RC & Woods RE, *Digital Image Processing*. World Student Series, ed. Addison-Wesley. 1993, Reading, Massachusetts: Addison-Wesley.
75. Fiorentini A, Maffe L, Sandini G, *The role of high spatial frequencies in face perception.* Perception, 1983, **12**: 195-201.
76. Näsänen R, *Spatial frequency bandwidth used in the recognition of facial images.* Vision Research, 1999, **39**: 3824-3833.

77. Gold J, Bennett PJ, Sekuler AB, *Identification of band-pass filtered letters and faces by human and ideal observers.* Vision Research, 1999, **39**: 3537-3560.
78. Hayes T, Morrone MC, Burr DC, *Recognition of positive and negative bandpass-filtered images.* Perception, 1986, **15**(5): 595-602.
79. Peli E, Lee E, Trempe CL, Buzney S, *Image enhancement for the visually impaired: The effects of enhancement on face recognition.* Journal of the Optical Society of America. A, Optics, image science, and vision, 1994, **11**(7): 1929-39.
80. Costen NP, Parker DM, Craw I, *Effects of high-pass and low-pass spatial filtering on face identification.* Perception and Psychophysics, 1996, **58**(4): 602-612.
81. Munhall KG, Kroos C, Jozan G, Vatikiotis-Bateson E, *Spatial frequency requirements for audiovisual speech perception.* Perception and Psychophysics, 2004, **66**(4): 574-583.
82. Nagayama R, Yoshida H, Toshima T, *Interrelationship between the facial expression and familiarity: analysis using spatial filtering and inverted presentation.* Shinrigaku Kenkyu, 1995, **66**(5): 327-335.
83. Nagayama R. *Personal communication*, 19th May 2006.
84. Schwartz O, Bayer HM, Pelli D, *Features, frequencies, and facial expressions.* Investigative Ophthalmology & Visual Science. ARVO abstract, 1998.
85. Schwartz GE. *Personal communication*, 20th May, 2006.
86. Vuilleumier P, Armony JL, Driver J, Dolan RJ, *Distinct spatial frequency sensitivities for processing faces and emotional expressions.* Nature Reviews Neuroscience, 2003, **6**(6): 624-31.
87. Bruce V & Valentine T, *When a nod's as good as a wink: The role of dynamic information in face recognition*, in *Practical Aspects of Memory: Current Research and Issues*, MM Gruneberg, PE Morris, and RN Sykes (ed.), 1988, John Wiley & Sons: Chichester.
88. Hill H & Johnston A, *Categorizing sex and identity from the biological motion of faces.* Current Biology, 2001, **11**(11): 880-885.
89. Knight B & Johnston A, *The role of movement in face recognition.* Visual cognition, 1997, **4**(3): 265-273.
90. Lander K, Christie F, Bruce V, *The role of movement in the recognition of famous faces.* Memory and Cognition, 1999, **27**(6): 974-85.
91. Lander K, Bruce V, Hill H, *Evaluating the effectiveness of pixelation and blurring on masking hte identity of familiar faces.* Applied Cognitive Psychology, 2001, **15**: 101-116.
92. Adolps R, Tranel D, Damasio AR, *Dissociable Neural Systems for Recognizing Emotions.* Brain and Cognition, 2003, **52**: 61-69.
93. Humphreys GW, Donnelly N, Riddoch MJ, *Expression is computed separately from facial identity, and it is computed separately for moving and static faces: Neuropsychological evidence.* Neuropsychologia, 1993, **31**(2): 173-181.
94. Bassili JN, *Facial motion in the perception of faces and of emotional expression.* Journal of Experimental Psychology: Human Perception and Performance, 1978, **4**(3): 373-379.
95. Harwood NK, Hall LJ, Shinkfield AJ, *Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation.* American Journal on Mental Retardation, 1999, **104**(3): 270-278.

96. Ambadar Z, Schooler JW, Cohn JF, *Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions.* Psychological Science, 2005, **16**(5): 403-410.
97. WHO, *International Classification of Diseases, 10th ed. (ICD-10).* in *Mental and behavioural disorders (chapter 5), diagnostic criteria for research.* 1993, World Health Organization (WHO): Geneva.
98. APA, *Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (DSM-IV).* 1994, American Psychiatric Association (APA): Washington DC.
99. Volkmar F, Chawarska K, Klin A, *Autism in infancy and early childhood.* Annual Review of Psychology, 2005, **56**: 315-336.
100. Barton JJS, Cherkasova MV, Hefter RL, Cox TA, O'Connor M, et al., *Are patients with social developmental disorders prosopagnotic? Percpetual heterogeneity in the Asperger and socio-emotional processing disorders.* Brain, 2004, **127**(8): 1691-2.
101. Hefter RL, Manoach DS, Barton JJS, *Perception of facial expressions and facial identity in subjects with social developmental disorders.* Neurology, 2005, **65**: 1620-1625.
102. Schultz R, *Developmental deficits in social perception in autism: the role of the amygdala and fusiform face are.* International Journal of Developmental Neuroscience, 2005, **23**: 125-141.
103. Dawson G, Webb SJ, McPartland J, *Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies.* Developmental Neuropsychology, 2005, **27**(3): 403-424.
104. Behrmann M, Avidan G, Leonard GL, Kimchi R, Luna B, et al., *Configural processing in autism and its relationship to face processing.* Neuropsychologia, 2006, **44**: 110-129.
105. Maurer D, Le Grand R, Mondloch CJ, *The many faces of configural processing.* Trends in Cognitive Science, 2002, **6**(6): 255-260.
106. Jemel B, Mottron L, Dawson M, *Impaired face processing in autism: Fact or artifact?* Journal of Autism and Developmental Disorders, 2006, **36**(1): 91-106.
107. Deruelle C, Rondan C, Gepner B, Tardif C, *Spatial frequency and face processing in children with autism and asperger syndrome.* Journal of Autism and Developmental Disorders, 2004, **34**(2): 199-210.
108. Barton JJS, Cherkasova MV, hefter R, Cos TA, O'Connor M, et al., *Are patients with social developmental disorders prosopagnosic? Perceptual heterogeneity in the Asperger and socio-emotional processing disorders.* Brain, 2004, **127**(8): 1706-1716.
109. Haxby JV, Hoffman EA, Gobbini MI, *The distributed human neural system for face perception.* Trends Cogn Sci, 2000, **4**(6): 223-233.
110. Tantam D, Monaghan L, Nicholson H, Stirling J, *Autistic children's ability to interpret faces: A research note.* Journal of Child Psychology and Psychiatry, 1989, **4**: 623-630.
111. Celani G, Battacchi MW, Arcidiacono L, *The understanding of the emotional meaning of facial expressions in people with autism.* Journal of Autism and Developmental Disorders, 1999, **29**(1): 57-66.

112. Gross TF, *The perception of four basic emotions in human and nonuman faces by children with autism and other developmental disabilities.* Journal of Abnormal Child Psychology, 2004, **32**(5).
113. Capps L, Yirmiya N, Sigman M, *Understanding of simple and complex emotions in non-retarded children with autism.* Journal of Child Psychology and Psychiatry, 1992, **33**(7): 1169-1182.
114. Castelli F, *Understanding emotions from standardized facial expressions in autism and normal development.* Journal of Autism and Developmental Disorders, 2005, **9**(4): 428-449.
115. Gepner B, Deruelle C, Grynfeltt S, *Motion and emotion: A novel approach to the study of face processing by young autistic children.* Journal of Autism and Developmental Disorders, 2001, **31**(1): 37-45.
116. Grossman JB, Klin A, Carter AS, Volkmar FR, *Verbal bias in recognition of facial emotions in children with Asperger Syndrome.* Journal of Child Psychology and Psychiatry, 2000, **41**(3): 369-379.
117. Baron-Cohen S, Spitz A, Cross P, *Do children with autism recognise surprise? A research note.* Cognition and Emotion, 1993, **7**(6): 507-516.
118. Adolphs R, Sears L, Piven J, *Abnormal processing of social information from faces in autism.* Journal of Cognitive Neuroscience, 2001, **13**(2): 232-240.
119. Baron-Cohen S, Jolliffe T, Mortimore C, Robertson M, *Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger Syndrome.* Journal of Child Psychology and Psychiatry, 1997, **38**: 813-822.
120. Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I, *The "Reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger Syndrome or high-functioning autism.* Journal of Child Psychology and Psychiatry, 2001, **42**: 241-252.
121. Baron-Cohen S, Wheelwright S, Jolliffe T, *Is there a "Language of the eyes"? Evidence from normal adults, and adults with autism or Asperger Syndrome.* Visual Cognition, 1997, **4**(3): 311-331.
122. Golan O, Baron-Cohen S, Hill J, *The Cambridge mindreading (CAM) face-voice battery: testing complex emotion recognition in adults with and without Asperger syndrom.* Journal of Autism and Developmental Disorders, 2006, **Epub ahead of print**.
123. Baron-Cohen S, Riviere A, Fukushima M, French D, Hadwin J, et al., *Reading the mind in the eyes: A cross-cultural and developmental study.* Visual Cognition, 1996, **3**(1): 39-59.
124. Neurobehavioral Systems Inc., *Presentation*, URL: http://www.neuro-bs.com/
125. Adolphs R, Tranel D, Damasio H, Damasio AR, *Impaired recognition of emotion in facial expressions following bilateral damate to the human amygdala.* Nature, 1994, **372**: 669-672.
126. Blair RJ, Morris JS, Frith CD, Perrett DI, Dolan RJ, *Dissociable neural responses to facial expressions of sadness and anger.* Brain and Cognition, 1999, **122**: 883-893.
127. Bagby RM, Parker JDA, Taylor GJ, *The twenty-item Toronto alexithymia scale.* Journal of Psychosomatic Research, 1994, **38**: 23-40.

128. Taylor GJ, Bagby RM, Parker JDA, *The alexithymia construct. A potential paradigm for psychosomatic medicine.* Psychosomatics, 1991, **32**: 153-164.
129. Salminen JK, Saarijärvi S, Äärelä E, Toikka T, Kauhanen J, *Prevalence of alexithymia and its association with sociodemographic variables in the general population of Finland.* Journal of Psychosomatic Research, 1999, **46**(1): 75-82.
130. Statsoft Inc., *Statistica for Windows*, URL: http://www.statsoft.com
131. Frydrych M, Kätsyri J, Dobsík M, Sams M. *Toolkit for animation of Finnish talking head*. in *AVSP*. 2003. St. Jorioz, France.
132. Faigin G, *The Artist's Complete Guide to Facial Expression*. 1990, New York: Watson-Guptill.
133. Nummenmaa T, *Pure and Blended Emotion in the Human Face*. 1992, Helsinki: Federation of Finnish Scientific Societies.
134. Tamminen T, Kätsyri J, Frydrych M, Lampinen J. *Joint modeling of facial expression and shape from video*. in *Scandinavian Conference on Image Analysis (SCIA)*. 2005.
135. Hager J & Ekman P, *The Inner and Outer Meanings of Facial Expressions*, in *Social Psychophysiology: A Sourcebook*, JT Cacioppo and RE Petty (ed.), 1983, The Guilford Press: New York.
136. Prevost S & Pelachaud C, *Talking Heads: Physical, linguistic and cognitive issues in facial animation*. Course Notes for Computer Graphics International. 1995, Leeds, UK.
137. Pelachaud C, Badler NI, Steedman M, *Linguistic issues in facial animation*, in *Computer Animation '91*, N Magnenat-Thalmann and D Thalmann (ed.), 1991, Springer-Verlag. pp 15-30.
138. Internet source: *Multimodal speech synthesis*. Department of speech, music and hearing, KTH, Stockholm. URL: http://www.speech.kth.se/multimodal/
139. Ostermann J. *Animation of Synthetic Faces in MPEG-4*. in *Computer Animation*. 1998. Philadelphia, Pennsylvania.
140. Massaro DW & Light J, *Using visible speech for training perception and production of speech for hard of hearing individuals.* Journal of Speech, Language, and Hearing Research, 2004, **47**(2): 304-320.
141. Internet source: *Ananova*. Orange Group. URL: http://www.ananova.com
142. Olives J-L, Möttönen R, Kulju J, Sams M. *Audio-visual speech synthesis for Finnish*. in *Auditory-Visual Speech Processing (AVSP'99)*. 1999. Santa Cruz, CA, USA.
143. Parke F & Waters K, *Computer Facial Animation*. 1996, Wellesley, Massachusetts: A K Peters.
144. Waters K, *A muscle model for animating three-dimensional facial expressions.* Computer Graphics, 1987, **21**(4): 17-24.
145. Badler NI & Platt S, *Animating facial expressions.* Computer Graphics, 1981, **13**(3): 245-252.
146. Terzopoulos D & Waters K, *Analysis and synthesis of facial image sequences using physical and anatomical models.* IEEE transactions on pattern analysis and machine intelligence, 1993, **15**(6): 569-579.
147. Beskow J. *Rule-based visual speech synthesis*. in *European Conference on Speech Communication and Technology*. 1995. Madrid, Spain.

148. Massaro DW, *Perceiving Talking Faces*. 1998, Cambridge, Massachusetts: The MIT Press.
149. Kalliomäki I & Lampinen J. *Feature-based inference of human head shapes*. in *Finnish Conference on Artificial Intelligence (STeP)*. 2002. Oulu, Finland.
150. O'Rourke M, *Principles of Three-Dimensional Computer Animation*. 1995, New York: W. W. Norton & Company.
151. Pighin F, Hecker J, Lischinski D, Szeliski R. *Synthesizing realistic facial expressions from photographs*. in *SIGGRAPH*. 1998.
152. Cohn JF, Ambadar Z, Ekman P, *Observer-based measurement of facial expression with the Facial Action Coding System*, in *The handbook of emotion elicitation and assessment*, JA Coan and JB Allen (ed.), In press, Oxford University Press Series in Affective Science: New York: Oxford.
153. Hess U, Blairy S, Kleck RE, *The intensity of emotional facial expressions and decoding accuracy*. Journal of Nonverbal Behavior, 1997, **21**(4): 241-257.
154. Lane RD, Sechrest L, Reidel R, Weldon V, Kaszniak A, et al., *Impaired verbal and nonverbal emotion recognition in alexithymia*. Psychosomatic Medicine, 1996, **58**: 203-210.
155. Matsumoto D & Ekman P, *American-Japanese cultural differnces in intensity ratings of facial expressions of emotion*. Motivation and Emotion, 1989, **13**(2): 143-157.
156. Kshirsagar S, Escher M, Sannier G, Magnenat-Thalmann N. *Multimodal animation system based on the MPEG-4 standard*. in *Multimedia Modeling*. 1999. Ontario, Canada.
157. Ahlberg J, Pandzic IS, You L. *Evaluating face models animated by MPEG-4 FAPS*. in *OZCHI, Talking Head Technology Workshop*. 2001. Perth, Western Australia.
158. Lavagetto F & Pockaj R, *The facial animation engine: toward a high-level interface for the design of MPEG-4 compliant animated faces*. IEEE Transactions on Circuits and Systems for Video Technology, 1999, **9**(2): 277-289.
159. Ahlberg J, *Model-based Coding: Extraction, Coding, and Evaluation of Face Model Parameters*. Doctor's Thesis. Department of Electrical Engineering, Linköping University.
160. Tekalp M & Ostermann J, *Face and 2d mesh animation in mpeg-4*. Image Communication Journal, 2000(Tutorial Issue on MPEG-4 Standard).
161. Lander K, Christie F, Bruce V, *The role of movement in the recognition of famous faces*. Mem Cognit, 1999, **27**(6): 974-85.
162. Ekman P, Friesen W, Hager J, *Facial Action Coding System*. 1978, Palo Alto, CA: Consulting Psychologists Press.
163. Kätsyri J, Klucharev V, Frydrych M, Sams M. *Identification of synthetic and natural emotional facial expressions*. in *ISCA Tutorial and Research Workshop on Audio Visual Speech Processing (AVSP)*. 2003. St. Jorioz, France.
164. Bassili J, N., *Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face*. Journal of Personality and Social Psychology, 1979, **37**(11): 2049-2058.
165. The MathWorks, *Matlab*, URL: http://www.mathworks.com

166. Tiippana K, Näsänen R, Rovamo J, *Contrast matching of two-dimensional compound gratings.* Vision Research, 1994, **34**(9): 1157-1163.
167. Poynton C, *Gamma and its disguises: The nonlinear mappings of intensity in perception, CRTs, film and video.* SMPTE Journal, 1993, **102**(12): 1099–1108.
168. Spencer J, O'Brien J, Riggs K, Braddick O, Atkinson J, et al., *Motion processing in autism: evidence for a dorsal stream deficiency.* Neuroreport, 2000, **11**(12): 2765-2767.
169. Bertone A, Mottron L, Jelenic P, Faubert J, *Motion perception in autism: A "complex" issue.* Journal of Cognitive Neuroscience, 2003, **15**(2): 218-225.
170. Duchaine BC & Nakayama K, *Developmental prosopagnosia and the Benton Facial Recognition Test.* Neurology, 2004, **62**: 1219-1220.
171. Korkman M, Kirk U, Kemp SL, *NEPSY.* 1997, Psykologien Kustannus OY: Helsinki.
172. Ehlers S, Gillberg C, Wing L, *A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children.* Journal of Autism and Developmental Disorders, 1999, **29**(2): 129-141.
173. Wechsler D, *Wechsler Adult Intelligence Scale - Revised (WAIS-R). Finnish translation.* 1981, Psykologien Kustannus OY: Helsinki.
174. Johnson MH, *Subcortical face processing.* Nature Reviews Neuroscience, 2005, **6**: 766-774.
175. Lander K & Bruce V, *Recognizing famous faces: Exploring the benefits of facial motion.* Ecological Psychology, 2000, **12**(4): 259-272.
176. Benson PJ & Perrett DI, *Synthesising continuous-tone caricatures.* Image and Vision Computing, 1991, **9**(2): 123-129.
177. Calder AJ, Rowland D, Young AW, Nimmo-Smith I, Keane J, et al., *Caricaturing facial expressions.* Cognition, 2000, **76**: 105-146.
178. LaBar KS, Crupain MJ, Voyvodic JT, McCarthy G, *Dynamic perception of facial affect and identity in the human brain.* Cerebral Cortex, 2003, **13**: 1023-1033.
179. Roark DA, Barrett SE, Spence MJ, Hervé A, O'Toole AJ, *Psychological and neural perspectives on the role of motion in face recognition.* Behavioral and Cognitive Neuroscience Reviews, 2003, **2**(1): 15-46.

Appendix A FACS meta-language

Expression	Matching action unit combinations	Examples
A	any action unit(s) with intensity from C to E; with/without additional expressions	1 ⇔ 1C to 1E all of the examples below
A*	any action unit with any intensity	1* ⇔ 1A to 1E
A*N*	any action unit with intensity/intensities *N*	1CD ⇔ 1C or 1D
[A]	empty or A	[1]+2 ⇔ 1 or 1+2
A_1|A_2	either A_1 or A_2	1|2 ⇔ 1 or 2
A_1/A_2	A_1, A_2 or both	1/2 ⇔ 1, 2 or 1+2
A_1|…|A_n	either A_1, …, A_{n-1} or A_n	1|2|4 ⇔ 1, 2 or 4
A_1/…/A_n	A_1, …, A_{n-1} or A_n or any of their combination	1/2/4 ⇔ 1, 2, 4, 1+2, 1+4, 2+4 or 1+2+4
A& or &A	A added to all of the following or preceding AU combinations	1& 2, 4, 5 ⇔ 1+2, 1+4, 1+5 1, 2, 4 &5 ⇔ 1+5, 2+5, 4+5
-A	Preceding AU combinations without A	1+2+4-1 ⇔ 2+4 1+2+4-1|2 ⇔ 2+4, 1+4

Table A.1 FACS meta-language expressions with explanations and examples.

A simple meta-language was devised for describing several sets of FACS (Facial Action Coding System) [1] action unit combinations in a compact form.

The meta-language contains expressions (Table A.1) extending the typical FACS notation [1]. Note that these expressions can be combined further with each other. A full meta-language expression refers to all action unit combinations matching it. Expressions are always evaluated from left to right with three exceptions: Expressions inside parentheses "()" precede other expressions, additive expressions "&" precede the remaining expressions and exclusions "-" are always evaluated last.

In typical FACS notation, action unit combinations are listed in an ascending order and separated by plus signs. For example, notation "1+2+4" refers to the simultaneous activation of action units 1, 2 and 4. Equivalently, when evaluating the meta-language expressions the action units are always kept in ascending order. Respectively, if in combination A+B either A or B is empty, the plus sign is omitted.

References

1. Ekman P, Friesen W, Hager J, *Facial Action Coding System*. 2nd ed. 2002, Salt Lake City: Research Nexus eBook.

Appendix B FACS prototypes for basic expressions

This appendix describes FACS (Facial Action Coding System) [1] action unit prototypes for basic emotions as suggested originally by the authors of FACS and as used in this thesis. Note that all prototypes are presented in a FACS meta-language notation (Appendix A) to allow for a compact presentation.

Table B.1 shows prototypes and their major variants [2: 173-174], and "critical actions" for basic emotions as suggested tentatively by the authors of FACS. Note that the authors have also suggested that various minor variants exist in addition to the major variants. Critical actions refer to facial actions used in EMFACS [2: 135-7, 3], a subset of FACS intended for coding only emotionally salient facial actions, as indicators of emotionally relevant events.

Emotion	**Prototypes**	**Major variants**	**Critical actions**
Anger	4+5*+7& [10*+[22]]+23+25|26,	Prototypes -4|5*|7|10	4+5/7, 17+24, 23
Disgust	9|10*+[16+25|26], 9|10+17		9, 10
Fear	1+2+4+5*+[20*]+25|26|27	1+2+4+5*+[L20|R20+25|26|27], (1+2+5DE, 5*+20* &[25|26|27])	1+2+4, 20
Happiness	6+12*, 12CD		[6|7]+12
Sadness	6+15*, 1+4+(11+15B)|15* &[25|26]+[54+64]	11+17, (1+4+11|(15B+[17]), 11+15B &[54+64]) &[25|26]	1+[4], [6]+15, 11+15|17
Surprise	1+2+5B+26|27	1+2+5B, 1+2+26|27, 5B+26|27	1+2+5AB/26

Table B.1 FACS action unit prototypes, their major variants and critical actions for coding emotion-related events as suggested by the authors of FACS.

Facial expression prototypes for six basic emotions, two of their blends and one non-emotional facial expression, intended as plausible examples rather than definitive models, are presented in Table B.2. The prototypes were designed by a certified FACS coder (JK) on the basis of existing literature [2: 135-7, 173-4, 4-6]. Action units were classified further into those primary and those secondary for the prototypes on the basis of [1, 2: 173-4, 4]. For all suggested prototypes, resembling action unit combinations with equal emotional interpretations were sought from FACSAID (FACS Affect Interpretation Dictionary) dictionary [7].

Emotion	Prototype(s)	Primary AUs	Secondary AUs	FACSAID equivalent
Anger	4+5+7+24	4+5+23|24	4+5+7, 17+23|24	4+5B+7+24
Disgust	9+10+17	9|10	9|10+17, 9|10+25	4+9B+10B+17B+61B+64A
Fear	1+2+4+5+7+20+25	1+2+4+5+7+20	20+23, 20+25	1+2+4+5B+7+20B
Happiness	6+12, 6+12+25	6+12	6+7+12, 12+[16]+25	6+12, 6+12+25 (Duchenne smile)
Sadness	1+4+7+15+17	1+7+15	1+4, 6+7, 15+17, 43	1+4C+7A+15B+17A+58+61C
Surprise	1+2+5+25+26|27	1+2+5+25+26|27	16+25+26+27	1+2+5B+26C
Happiness & Surprise	1+2+5+6+12+25+26|27	*Unspecified*		1+2+5B+6+12B+50 (Surprise/Duchenne smile)
Happiness & Disgust	6+9+10+12+17			R4B+R6+R9B+10B+12B+17B (Disgust/Possible Duchenne smile)
Mouth opening	25+26			26 (No prediction)

Table B.2 Hypothetical FACS prototypes, primary and secondary action units (AUs) and equivalent action unit combinations from FACSAID emotional facial expression dictionary for six basic emotions, two blended emotions and one non-emotional facial expression. Within each secondary action unit combination, the underlined item refers to the main action and the non-underlined items to its necessary context (*e.g.* with anger, AU7 is considered secondary only in combination with AU4+5).

References

1. Ekman P, Friesen W, Hager J, *Facial Action Coding System*. 2nd ed. 2002, Salt Lake City: Research Nexus eBook.
2. Ekman P, Friesen W, Hager J, *Facial Action Coding System: Investigator's Guide*. 2nd ed. 2002, Salt Lake City: Research Nexus eBook.
3. Ekman P, Irwin W, Rosenberg EL, *EMFACS-7*. 1994.
4. Ekman P, Friesen W, *Unmasking the face. A guide to recognizing emotions from facial expressions*. 1975, Palo Alto, CA: Consulting Psychologists Press.
5. Faigin G, *The Artist's Complete Guide to Facial Expression*. 1990, New York: Watson-Guptill.
6. Nummenmaa T, *Pure and Blended Emotion in the Human Face*. 1992, Helsinki: Federation of Finnish Scientific Societies.
7. Internet source: *DataFace*. http://face-and-emotion.com/dataface

Appendix C FACS evaluation of TKK collection

Actor	Emt	FACS evaluation		
		JK	**VK**	**Agreem.**
KH	ang	4C+5D+6A+7C+10A+23C+24C+31B+38C	4C+5B+7B+10A+24C	83 %
KH	dis	9D+10D+17B+26	9D+10B+17B	86 %
KH	fea	1C+2C+4B+5D+10C+11D+20C+25	1B+2D+4B+5D+10D+11D+L15A+20C+25C	94 %
KH	hapC	6D+12D+24C		
KH	hap	6C+12D+25	6C+12D+25C	100 %
KH	sad	1B+2B+4B+7C+10A+15B+17B	1B+2B+4B+7B+15B+39B	83 %
KH	sur	1D+2D+5C+25+26	1E+2E+5B+25C+26D+38C	100 %
ME	ang	4C+5D+7B+17D+24C		
ME	dis	9D+10B+17C+24B		
ME	fea	1B+2B+4A+5D+7B+20C+25		
ME	hapC	6E+12D+17C		
ME	hap	6C+12D+R16B+25		
ME	sad	1B+4A+7B+12B+17C+20C+38C+43C		
ME	sur	1C+2C+5D+25+27		
MR	ang	4B+5C+7C+24B+38C		
MR	dis	4B+7B+9C+10C+17A		
MR	fea	1C+2C+4B+5C+20A+T23B+25+38B		
MR	hapC	6B+13D		
MR	hap	6B+12D+25		
MR	sad	1B+4C+7B+15B+17B+20B		
MR	sur	1C+2C+5C+25+27+38B		
NR	ang	4E+5D+6B+7C+17D+23C+24C	4D+5A+7C+9A+23B+24C	77 %
NR	dis	4C+7B+9D+10E+17D	4C+7B+9D+17B	89 %
NR	fea	1C+2C+4A+5D+20D+25	1C+2B+5D+16B+20C+25C	83 %
NR	hapC	6D+12C+17C		
NR	hap	6C+12D+16B+25	6B+12D+25B	86 %
NR	sad	1B+4B+15B+17D+39B	1A+4B+15A+17C+39C	100 %
NR	sur	1C+2C+5D+16B+25+26	1C+2C+5D+25C+26E	91 %
SP	ang	4E+5D+7B+10A+17B+24C	4D+5B+10A+24C	80 %
SP	dis	4D+6B+7A+9C+10C+15B+17D	4C+9D+10C+17C	73 %
SP	fea	1D+2D+5B+12B+20B+25	1D+2C+5A+12B+20C+25B+38C	100 %
SP	hapC	6C+10A+12C		
SP	hap	6D+12D+16B+25	6B+12C+25D	86 %
SP	sad	4A+6B+7D+15B+17C+20B	6B+7D+15B+17B	80 %
SP	sur	1E+2E+5C+25+27+38C	1D+2D+5B+25A+27D	100 %
TV	ang	4B+5C+6B+7D+15B+23D+24A+38D	4A+5B+7C+23C+38B	73 %
TV	dis	6B+7D+9E+10C+17C	7D+9C+24A	50 %
TV	fea	1C+2B+4B+5C+7C+20E+21D+25+38C	1B+4C+5C+7A+20D+25C+38D	92 %
TV	hapC	6E+7B+12E+16A		
TV	hap	6D+12D+25	6C+12D+25D	100 %
TV	sad	1C+4B+6D+7E+10A+15E+17E	4B+7E+15D+17C+38C	73 %
TV	sur	1E+2E+5D+25+27+38B	1B+2D+5C+25C+26E+38C	80 %

Table C.1 FACS coding for TKK collection. Columns from left to right are: actor initials, posed emotion (three first letters; hapC denotes closed-mouth happiness variant), FACS evaluation by two coders (JK and VK) and agreement evaluation between them.

Actor	Emt.	Prototypical action units		
		Primary	**Missing**	**Extra**
KH	ang	4C+5D+24C		6A+10A
KH	dis	9D		26
KH	fea	1C+2C+4B+5D+20C	7	10C+11D
KH	hapC	6D+12D		24C
KH	hap	6C+12D		
KH	sad	1B+4B+7C+15B		2B+10A
KH	sur	1D+2D+5C+25+26		
ME	ang	4C+5D+24C		
ME	dis	9D		24B
ME	fea	1B+2B+4A+5D+20C	7	
ME	hapC	6E+12D		17C
ME	hap	6C+12D		
ME	sad	1B+4A+7B	15	12B+20C
ME	sur	1C+2C+5D+25+27		
MR	ang	4B+5C+24B		
MR	dis	9C		4B+7B
MR	fea	1C+2C+4B+5C+20A	7	
MR	hapC	6B	12	13D
MR	hap	6B+12D		
MR	sad	1B+4C+7B+15B		20B
MR	sur	1C+2C+5C+25+27		
NR	ang	4E+5D+24C		6B
NR	dis	10E		4C+7B
NR	fea	1C+2C+4A+5D+20D	7	
NR	hapC	6D+12C		17C
NR	hap	6C+12D		
NR	sad	1B+4B+15B	7	
NR	sur	1C+2C+5D+25+26		
SP	ang	4E+5D+24C		10B+17B
SP	dis	9C		4D+6B+7A+15B
SP	fea	1D+2D+5B+20B	4+7	12B
SP	hapC	6C+12C		10A
SP	hap	6D+12D		
SP	sad	4A+7D+15B	1	6B+20B
SP	sur	1D+2D+5C+25+27		
TV	ang	4B+5C+23D		6B+15B
TV	dis	9E		6B+7D
TV	fea	1C+2B+4B+7C+20E	5	
TV	hapC	6E+12E		7B+16A
TV	hap	6D+12D		
TV	sad	1C+4B+7E+15E		6D+10A
TV	sur	1E+2D+5D+25+27		

Table C.2 The FACS evaluation of TKK stimuli classified on the basis of their prototypical facial actions.

FACS [1] coding for basic emotion stimuli in TKK collection by two certified FACS coders (JK and VK) is presented in Table C.1. Note that all stimuli haven't been evaluated by the second FACS coder (VK) because his evaluation was used mainly to confirm the validity of the first coder's evaluations. Agreement evaluation between the two coders for a particular stimulus has been calculated as twice the sum of agreed action units divided by the sum of all coded action units, resulting in a proportion between 0-100 %. In Table C.2, the action units evaluated by the first FACS coder have been classified into existing and missing primary prototypical facial actions and extra facial actions in addition to primary and secondary actions (cf. Appendix B for the definitions of primary and secondary facial actions).

References

1. Ekman P, Friesen W, Hager J, *Facial Action Coding System*. 2nd ed. 2002, Salt Lake City: Research Nexus eBook.

Appendix D TKK talking head parameters

Action unit	Name	Param.'s	Areas	Dir.	Secondary actions
AU1	Inner brow raiser	4	Le/Ri inner brow, Le/Ri inner eye cover	V	
AU2	Outer brow raiser	4	Le/Ri outer brow, Le/Ri outer eye cover	V	
AU4	Brow lowerer	2	Le/Ri brow	V	
AU5/41	Upper lid raiser/ Eye closer	2	Le/Ri Up eyelid	V	
AU6	Cheek raiser	2	Le/Ri cheek	V	- AU2, AU10 (L/R only)
AU7	Lid tightener	2	Le/Ri Lo eyelid	V	
AU9	Nose wrinkler	4	Le/Ri Up and Lo nasal areas	V	
AU10	Upper lip raiser	3	Le/Ri/Mi Up lip	O/O/V	+AU38
AU11	Nasolabial furrow deepener	2	Le/Ri nasolabial furrow area	O	+AU10 (L/R only)
AU12	Lip corner puller	2	Le/Ri lip corners	O	+AU11
AU15	Lip corner depressor	2	Le/Ri lip corners	O	
AU16/17	Lower lip depressor/ Chin raiser	3	Le/Ri Lo lip, Mi Lo lip/chin	O/O/V	-AU25+26/27
AU20	Lip stretcher	2	Le/Ri lip area	H	
AU23/24	Lip tightener/presser	6	Le/Ri/Mi Up and Lo lip areas	R	-AU20
AU25+26/27	Lips part/Jaw drop/ Mouth stretch	1	Jaw	V	
AU38/39	Nostril dilator/ compressor	2	Nostrils	R	

Table D.1 FACS [1] action unit modeling in TKK talking head [2]. Columns from left to right are: FACS action unit identifier(s), name of the action unit(s), number of parameters used to model the action unit(s), main areas affected by the parameters (Le left, Ri right, Lo lower and Up upper), movement direction (V vertical, H horizontal, O oblique and R orbital) and secondary actions caused by the activation of the action unit(s).

References

1. Ekman P, Friesen W, Hager J, *Facial Action Coding System*. 2nd ed. 2002, Salt Lake City: Research Nexus eBook.
2. Frydrych M, Kätsyri J, Dobsík M, Sams M, *Toolkit for animation of Finnish talking head*. in *AVSP*. 2003. St. Jorioz, France.