The Reconstruction of Facial Expressions in Embodied Systems
New Approaches to an Old Problem

Karl Grammer & Elisabeth Oberzaucher

***Karl Grammer** studied Zoology, Physics and Anthropology at the University of Munich. He received his PhD in Biology in 1982 at the University of Munich and the Research Institute for Human Ethology, Max-Planck-Society. His PhD thesis deals with self organization in groups of preschool children. In 1991 he became the Scientific Director of the Ludwig-Boltzmann-Institute for Urban Ethology in Vienna and was appointed Professor by the University of Vienna. In 2002 he received the Zdenek-Klein Award for his integrative scientifc work and 2003 he received the first prize in the International competition Vienna-Cooperate for outstanding scientific and technological innovation. Currently he is working on communication research and non-verbal behavior simulation in embodied agents at the Center for Interdisciplinary Studies (ZiF) in Bielefeld/Germany as a resident fellow.*

***Elisabeth Oberzaucher** obtained her MA in Zoology and her doctorate in Anthropology at the University of Vienna. Her focus of research lies in non-verbal behavior in social interactions. Social interactions range from zero-acquaintance situations to mate choice and social networks, especially friendships. Nonverbal behavior is of interest from an evolutionary viewpoint, since there exist adaptations to meet the need to make predictions of the other's behavior in social interactions. The current main research project is on the functionality of body odor. She is currently holding the position of interim director of the Ludwig-Boltzmann-Institute for Urban Ethology.*

Zusammenfassung
In der Ausdrucks- und Emotionsforschung lassen sich viele Beschränkungen auf das Fehlen eines geeigneten Stimulimodells zurückführen. Im vorliegenden Artikel wird ein neues, dreidimensionales Stimulimodell für die Erforschung des Gesichtsausdrucks vorgestellt. Es basiert auf dem Facial Action Coding System *von Paul Ekman & Wallace Friesen (1978), einer Methode, die entwickelt wurde, um jede Gesichtsbewegung anhand der zugrunde liegenden Muskelkontraktionen beschreiben zu können. Entsprechend kann man mit dem neuen Stimulimodell so gut wie jede Gesichtsmuskelkontraktion simulieren. Damit ist es nun möglich, die Intensität des Ausdrucks systematisch zu variieren, neue Ausdrücke zu generieren und Mischemotionen sowie die Dynamik von Gesichtsausdrücken unter konstanten Bedingungen zu untersuchen und zu simulieren.*
*Vor dem Hintergund verschiedener Emotionstheorien geht der Artikel speziell auf die Bedeutung der Komponenten eines Gesichtsausdrucks und deren Implikation für ein System zur Computersimulation von Gesichtsausdrücken ein. Einige Forscher vertreten die Ansicht, dass die einzelnen Komponenten (Muskelbewegungen) für sich alleine genommen bedeutungslos sind; erst in der Kombination erlangen sie Bedeutung. Demnach wäre der Gesichtsausdruck als Ganzes die Grundeinheit der Bedeutung. Andere Forscher vertreten dagegen die Ansicht, dass die Information bereits auf der Ebene der einzelnen Komponenten des Ausdrucks verschlüsselt ist. Demnach käme also die Bedeutung des Gesamtausdrucks durch die Bewertung der Komponenten anhand zugrunde liegender Dimensionen wie z. B. Erregung (*arousal*) und Valenz (*pleasure*) zustande.*
In der beschriebenen Studie haben Versuchspersonen per Zufall generierte komplexe (d. h. aus mehreren Komponenten / Aktionseinheiten bestehende) Gesichtsausdrücke in Hinsicht auf die Dimensionen ›Valenz‹ und ›Erregung‹ bewertet. Aus den Bewertungen wurden Regressionsflächen (mit den Variablen ›Intensität der Ausdruckskomponente‹, ›Valenz-‹ und ›Erregungsbewertung‹) für jede der verwendeten Muskelbewegungen berechnet. Die meisten dieser Ausdruckskomponenten zeigen eine spezifische Regressionsfläche, also eine Oberfläche, die nur ihnen eigen ist. Dies beweist, dass jede dieser Komponenten, abhängig von ihrer Intensität, tatsächlich eine eigene Bedeutung in Hinsicht auf die Dimensionen ›Erregung‹ und ›Valenz‹ besitzt. Diese Information kann nun zum Bau einfacher Computer- Kontrollarchitekturen verwendet werden, mit denen sich komplexe Gesichtsausdrücke simulieren lassen.

Human-machine interaction: Do we need emotional computers?

In their book *The Media Equation* REEVES and NASS (1996) demonstrate that people interact with computers like with real people. This may also indicate that people themselves prefer to be treated by computers in an emotional way. But what should an interaction between man and machine look like in order to meet social and emotional demands of humans, and to be comparable to real life situations? One way is the implementation of emotional feedback from the computer via non-verbal behavior, because people react most intensely to nonverbal communication, even if it is abstract (REEVES & NASS, 1996).

There are various attempts to model adaptive behavior in autonomous agents (STALLER & PETTA, 1998) which are situated in highly complex dynamic environments. One of the aims is to integrate a computational model of emotional processes within the architecture of embodied systems (interactive virtual characters). In this approach emotional systems are considered as systems that continuously monitor the relationship of an individual to its environment and instigates tendencies to act. These virtual characters are usually used for interactive kiosk applications, user guides for complex machines or computers, feedback in learning software, therapy, and internet communication (CASSELL, 2004). Although the implementation of emotion processes in such systems is highly sophisticated none of these projects offer a convincing solution for the action expression problem (i.e. mapping behavioral output on emotions). For example, PETTA et al. (1999) describe the construction of such a system in a virtual agent, but they do not provide a rationale how the emotional output and the expressive behavior of their agent are linked. The expressive behavior of this agent is based on pre-captured motions and surface color changes. In their outline of the emotional architecture STALLER and PETTA (1999) linked 24 emotion categories to 14 action response categories with over a thousand individual actions. It is unclear though, how the categories and actions were connected to each other. Most existing attempts to incorporate emotional signals in embodied systems are library solutions, i.e. they map emotions on expressive behaviors one to one. This mapping is usually done by hand. It is obvious that such systems are not flexible and tend to be perceived as monotonous. Thus, these restricted systems are not sufficient to lead to the desired effect in perception, because in social interactions signals tend to be more complex.

Here we present an approach we call *reverse engineering*, in accord with the technical term that describes the process of taking apart an object to see how it works in order to duplicate or enhance it. How can this be applied to the study of emotions?

Basically, any expression simulation system consists of two parts. One part is the control architecture and another part is the expressive output. The control architecture itself has to decide what facial expression has to be shown under which circumstances. In terms of engineering this control architecture must include variables, which are used to bring up facial expressions under defined circumstances. Thus the construction of such architecture needs to deal with the linkage between emotion theories and facial expressions. The other part seems to be resolved much easier—faces are equipped with muscles and muscles deform the surface of a face to subsequently form patterns of facial expressions. Further, the expressive system construction can rely either on top-down constructed expressive templates (whole expressions) or a bottom-up approach where singles muscles are animated (see below).

In the present study we tested the quality of such an emotional expressive system by segmenting it into its parts (according to the reverse engineering). On this basis we finally reconstructed a new system by means of simulation and tested its performance with real people. However, before we turn into the details of this *reverse engineering approach* we need to discuss some of the very basics of emotion and facial expression theories.

Constructing control architecture: facial expressions

Various attempts to define emotion have not let to a satisfactory result. The approach by PAUL EKMAN (1984) was probably the most successful attempt to an empirical definition of emotion: Emotions last for five-hundred milliseconds up to four seconds and are elicited by universal triggers. These triggers can be defined for different emotions, such as loss for sadness, unexpected events for surprise etc. and undergo a subjective non-conscious process called appraisal. Appraisal theories of emotion hypothesize that emotions are the result of a meaning analysis in which an individual evaluates the personal significance of a given stimulus occurring in the environment (LAZARUS, 1991). The appraisal of a situation generates a certain emotional state that evokes specific actions corresponding to this situation. Emotions are internal states that can be expressed by certain components of facial expressions. Activation of these components can intensify and modify communication, but although they may suit other purposes, all facial expressions result from emotions in the first place. Appraisal theories have already been modeled by WEHRLE and SCHERER (2003) and incorporated in emotion simulations in artificial intelligence (e.g. STALLER & PETTA, 1999), but to our knowledge the relation between control architecture and facial expression has not yet been tackled on an empirical level.

TOMKINS (1962), EKMAN (1971) and IZARD (1971) proposed the theory of a *Facial Expression Program* and suggested the connection between internal physiological processes, emotions and facial expression. When we experience an emotion, a cascade of electrical impulses (arising from the emotion centres of the brain) triggers a special facial expression, and certain physiological changes, such as increased or decreased heart rates, or changes in blood pressure. EKMAN defines the so-called *base emotions* as those emotions that are always linked to a specific facial expression. This linkage is hardwired (genetically determined) and all other facial expressions can be regarded as the result of mixtures of base emotions. This approach has gained support from neurophysiology (e.g. PANKSEPP, 1992, identifies the brain circuits that would correspond to these basic emotions). Emotions are regarded as physiological adaptations to external situations, facilitating the predictability of actions and reactions. Hence, facial expressions are the external signals of these emotions and we assume that it has been evolutionarily advantageous to signal one's internal state to the environment. Ekman (1980) describes the appraisal as an automatic mechanism which is a cognitive subsystem operating independently from other cognitive systems, that is dedicated to determine whether a stimulus will elicit a basic emotion (joy, fear, surprise, anxiety, disgust, anger, contempt).

In contrast to this view the '*Componential Approach to Emotions'* (FRIJDA, 1986; LANG, 1995; LAZARUS, 1991; SCHERER, 1984) sees emotions as structures consisting of certain, correlated components: The cognitive appraisal, physiological arousal, motoric system activation, subjective feeling and the motivational system (action disposition). Experiencing emotions is regarded as becoming conscious of one of these components. Facial expressions then reflect how an individual is dealing with his environment and how this individual is positioned in his environment: Acceptance or rejection, approach or avoidance, level of activation, and quality of activation (e.g. free from / or influenced by fear or nervousness). The componential process is a response to the evaluation of an external or internal stimulus event on a non-accessible cognitive level. During cognitive appraisal an individual constantly evaluates the ongoing situation according to relevance (novelty check, intrinsic pleasantness, goal relevance check), implications (causal attribution check, outcome probability check, discrepancy from expectation check, goal / need conduciveness check, urgency check), coping potential (control check, power check, adjustment check) and normative significance (internal and external standard check). In SCHERER's (1999, 2001) terms the resulting patterns of appraisal are associated with specific emotions, like joy, fear, sadness or anger. As different subsystems of appraisal might bring different evaluative results, the resulting emotion can be a mixture of the basic emotion dimensions above.

Facial expressions then mirror the states of action readiness, in other words, the willingness to interact, and can help an interacting person to adjust his behavior to the sender's current state. But facial expressions can also be used consciously in order to comment on observed situations and other persons' actions. The connection between facial expressions and emotions is rather loose, as they can occur independently.

All approaches cited above get into difficulties when it comes to the mapping of actual behavior onto emotional states. The discrete approach suffers from the fact that basic emotions are not represented by only one facial expression for each basic emotion, instead basic emotions form expression families and it is unclear on which specific pattern the result of an appraisal should be mapped. The componential approach also does not have specific rules for mapping emotions to facial expressions—it is rather vague when it comes to this problem.

RUSSELL (1995) rejects the categorical and the componential model of emotions, stating that emotions can hardly ever be observed in real life situations, but rather occur as “melodramatic poses” in conversations, underlining the spoken words. RUSSELL claims that the face does not convey more information about emotions than the rest of the body (e.g. body posture, words, intonation), but rather signals the global feelings of a person. He had found earlier (1991) that participants rating faces could only assess the level of arousal, and if the expressed feelings were positive or negative, but not interpret specific emotions. Only when the context was known, specific emotions could be recognized.

According to RUSSELL facial expressions contain primary information that can be automatically, quickly, and universally understood: Quasi-physical information, such as muscular contractions, skin color, tears, sweat, etc. on one hand, and the dimensions pleasure (pleasure-displeasure) and arousal (arousal-sleep) on the other hand. This primary information is associated with the context of the situation and allows the attribution of an emotion.

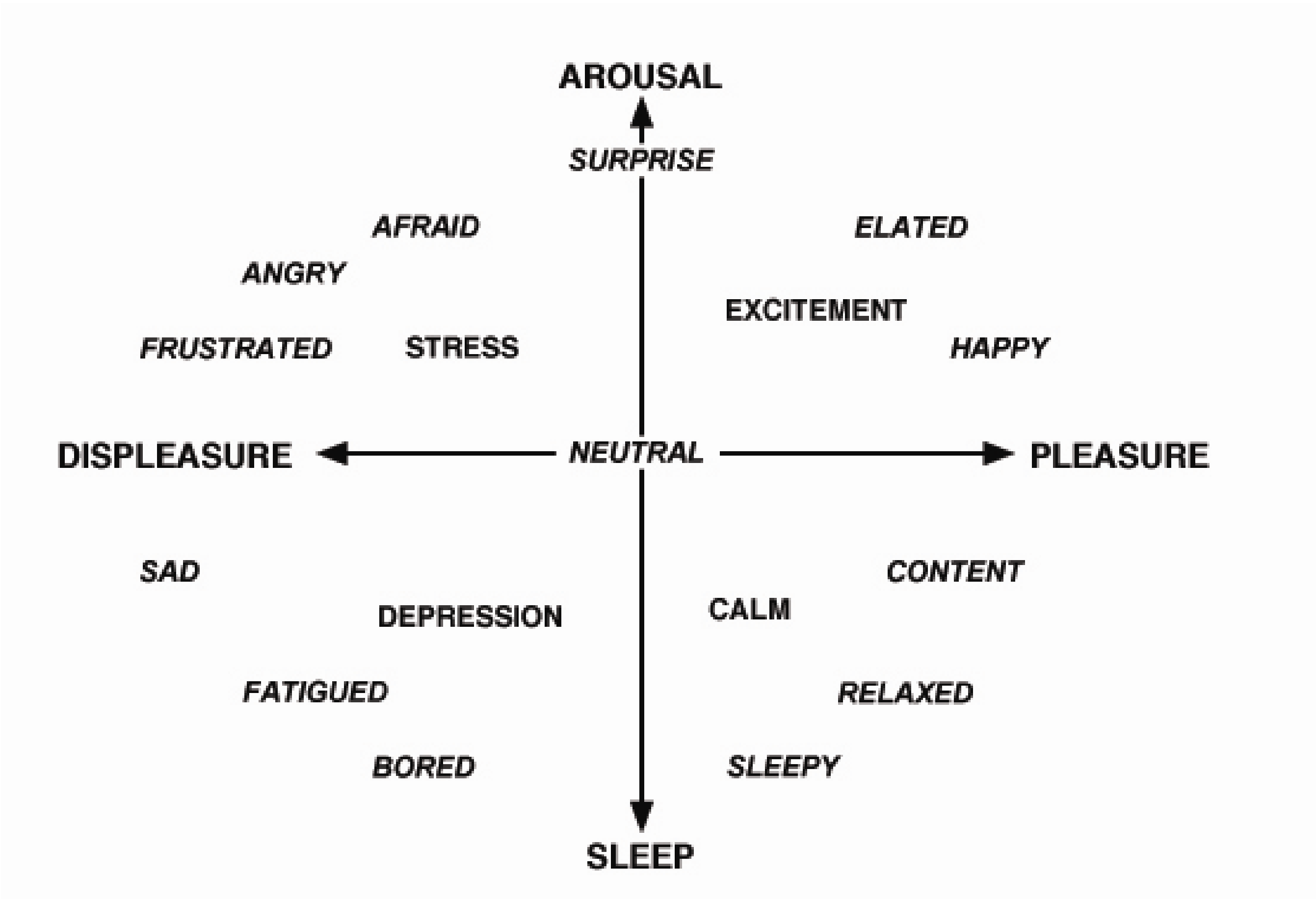

Figure 1: The *circumplex model* of emotions—the dimensions reach from displeasure to pleasure (or valence) and from sleep to arousal. Basic emotions are distributed at specific point within the pleasure and arousal space.

In the resulting *Circumplex Model* (RUSSELL, 1977) emotional states can be collocated in a two-dimensional space along the axes arousal-sleep and displeasure-pleasure. RUSSELL (1978) claims that two axes are sufficient, as further dimensions can only be components of some, but not of all models of emotions. Specific values of pleasure and arousal can be assigned to all emotions and facial expressions. They are collocated circularly around the “neutral” point, where the values for arousal and pleasure are at zero (*Circumplex Model*).

In contrast to the discrete and componential approach the circumplex model predicts a direct mapping of pleasure and arousal on each facial muscle, which is innate and hardwired. This allows us to use a non-interpretative bottom-up approach in the research on facial expression, which is independent from emotion categories.

Up to now most studies on the meaning of facial expressions followed a top-down approach: A certain facial expression was presented and an emotion term had to be attributed. Mostly, only a limited number of emotions were provided as possible answers. This method assumes that concordance in the attributions to a certain facial expression allows the conclusion that the attributed emotion is the reason for this facial expression. A bottom-up approach would instead use single facial movements (i.e. changes caused by a single muscle), which comprise a facial expression. Thus any interpretation on the level of emotion categories would be avoided. SNODGRASS (1992) was one of the first researchers to apply this bottom-up approach: Single muscle components were rated by observers according to their degree of pleasure and arousal. In a second step, they were attributed to emotion terms. In both settings subjects showed a high agreement in their ratings. Additionally, the attributed emotion was associated with the values in the pleasure and arousal dimensions. These findings support the assumption of RUSSELL (1980), who stated that the information conveyed by a facial expression is present even in the single components. The great advantage of this bottom-up approach is the possibility to investigate any type of expressive behavior and not only a limited set of emotions.

Many—and often contradictory—theories on the association between emotions and facial expressions make the decision for appropriate control architecture an almost impossible task. In this article we propose a new approach how to generate control architectures and the respective mapping on facial expressions. The *reverse engineering* look at the existing theories reveals that all use at least some overlapping building blocks. At first every theory needs some appraisal process in order to elicit an emotion. The result of this appraisal process is then being fed into a base module, which calculates the respective variables for the emotion elicitation process. At this point, any control architecture needs some expert system to decide which emotion is going to occur. After the emotion process is running, a facial expression occurs or not, again based on an expert system. Basically all three main theoretical approaches, the expression program, the componential approach and the pleasure and arousal *Circumplex Model* can be modeled this way. The basic problem for all theoretical approaches is the fact that an expert system is needed for translating appraisal into emotions and emotions into facial expression.

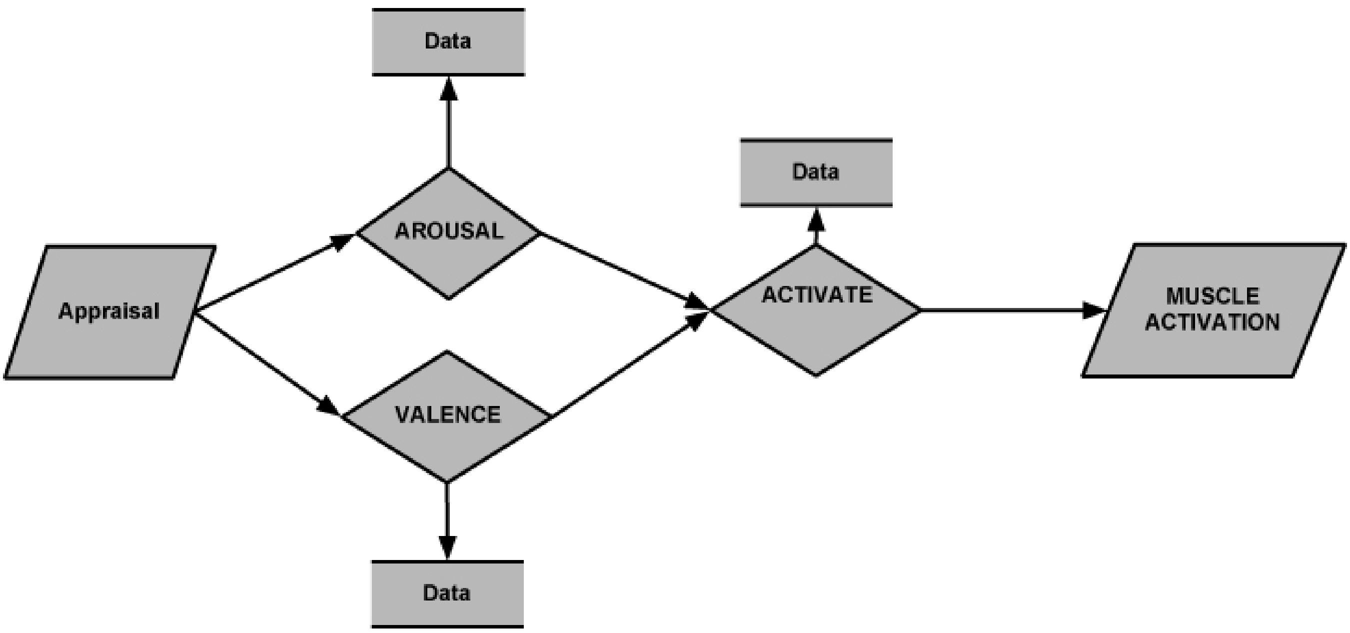

Figure 2: Control architecture for the *circumplex model*. This model shows the flow from perception of the environment to muscle activation in dependency form pleasure and arousal.

Figure 2 shows the flowchart for the possible implementation of a control architecture for the *Circumplex Model*. Again, an appraisal process (which is not discussed here) elicits two processes: An arousal and a valence process. The basic dimensions of appraisal can be mapped easily on pleasure and arousal. These values can then be directly mapped on the facial muscles in an expert system.

Up to now research on emotional expressions was executed using photographs of real people. These photos usually show the basic emotions in very high intensity. The shortcomings of this method are obvious: There is no possibility to exclude personal information (such as distinct physical characteristics) from the stimuli. Additionally, it is not possible to vary the intensity of the expressions systematically. Therefore expressions of low intensity have not been investigated so far. This is even more regrettable since we experience in our everyday life that we are capable of understanding subtle facial expressions just as well as ones of high intensity. One possibility to overcome these problems is the development of realistic 3d representations of faces.

The computer as device for expression research

The first face models on the computer were developed in the 1970ies. PARKE (1972) developed a parametric three-dimensional model and GILLENSON (1974) created the first interactive two-dimensional model. PLATT and BADLER (1981) designed the first model based on muscular actions, using FACS (Facial Action Coding System) as basis for the control of facial expressions.

FACS is the most widely used and versatile method for measuring facial behavior. PAUL EKMAN and WALLACE FRIESEN (1978) developed FACS by determining how the contraction of each facial muscle (singly and in combination with other muscles) changes the appearance of the face.

FACS has become most popular in emotion research. Action Units—the core elements of this system—are defined as surface structures that emerge due to contractions of single or groups of facial muscles that can be contracted independent from other facial muscles. The action units are labeled value-free with numbers and described very thoroughly in their special quality. The great advantage of FACS is that all possible facial changes respectively movements can be recorded and catalogued. It is strictly descriptive and has no need to refer to emotions.

Based on FACS SPENCER-SMITH et al. (2001) developed a first realistic three-dimensional model; it allows creating stimuli with 16 different action units and defined intensities. The limitations of this lie primarily in the number of embedded action units, taking into account that FACS consists of 46 action units for facial expressions and 12 action units for head and eye movements. Moreover, the base character of this model had rather few polygons with the result that the simulated facial expressions were rather crude approximations and details were lost.

Implementing FACS on a three dimensional computer face

We used 3D faces from the program “POSER 4” (Curious Labs, Santa Cruz, CA) as a test bed for our approach. Such a software environment brings the advantage that the lighting conditions, camera angle and focus length are exactly the same for all pictures. The avatars are three dimensional mesh models, consisting of single polygons, which can be textured arbitrarily according to the desired looks of the resulting figure. By adding hair models and photo realistic high-resolution textures realistic models can be created. We used the model “Victoria V.2” from “DAZ” (Digital Art Zone, Draper, UT) with a polygon count of 10434 for the head. We modeled all AUs from FACS as a system of morph targets at their maximum contraction directly on the head mesh in a modeling program. The meshes were translated by scientists and constantly rechecked by trained FACS coders. 25 Action Units were implemented this way and finally tested for interactions with codes from the FACS handbook.

By activating the single morphs of action units in almost any intensity and combination almost any type of facial expression can be constructed. Additionally, if a random element is added to the activation, two emotional expressions will be never the same. Thus, the major problem of facial animation—sterility through repetition—is eliminated.

AU	Definition	Involved muscles
AU 1	Inner Brow Raiser	M.frontalis, Pars medialis
AU 2	Outer Brow Raiser	M.frontalis, Pars lateralis
		M.depressor glabellae
		M.depressor supercilii
AU 4	Brow Lowerer	M.corrugator supercilii
AU 5	Upper Lid Raiser	M.levator palpebrae superioris
AU 6	Cheek Raise	M.orbicularis oculi, Pars orbitalis
AU 7	Lids Tight	M.orbicularis oculi, Pars palpebralis
AU 9	Nose Wrinkler	M.levator labii superioris alaeque nasi
AU 10	Upper Lip Raise	M.levator labii superioris, Caput infraorbitale
AU 11	Nasolabial Deepener	M.zygomaticus minor
AU 12	Lip Corner Pull	M.zygomaticus major
AU 13	Cheek Puff	M.levator anguli oris/ M.caninus
AU 14	Dimpler	M.buccinator
AU 15	Lip Corner Depressor	M.depressor anguli oris
AU 16	Lower Lip Depressor	M.depressor labii
AU 17	Chin Raise	M.mentalis
		Mm.incisivi labii superioris
AU 18	Lip Pucker	Mm.incisivi labii inferioris
AU 20	Lip Stretch	M.risorius
AU 22	Lip Funneler	M.orbicularis oris
AU 23	Lips Tightener	M.orbicularis oris
AU 24	Lips Pressor	M.orbicularis oris
		M.masseter
AU 26	Jaw Drop	Mm.pterygoidei (relaxed!)
		Mm.pterygoidei
AU 27	Mouth Stretch	M.digastricus
AU 38	Nostril Dilatator	M.nasalis, Pars alaris
		M.nasalis, Pars transversa
AU 39	Nostril Compressor	M.depressor septi nasi
AU 43	Eyes Closed	M. levator palpebrae superioris (relaxed!)

Table 1: FACS Action Units implemented in the system

Methods of stimulus generation and presentation

With these developments we were able to use a radical bottom-up approach, which was previously not possible in expression research. We developed a program called *Face Randomizer*, which allowed to randomly altering the intensities of the activations of single morph targets.

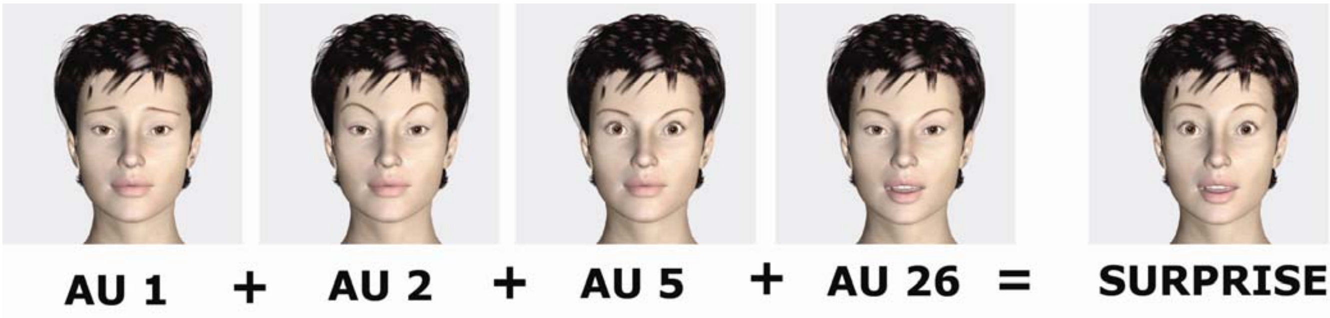

Figure 3: This picture shows how single modeled facial action units from the Facial Action Coding System can add up to a complex expressive facial pattern (AU1 Inner Brow Raise, AU2 Outer Brow Rise, AU5 Upper Lid Raise, AU26 Jaw Drop).

In the present study we used three different types of facial expressions, which were randomly presented to subjects:

- Type 1: Only one action unit is activated
- Type 2: Fifty percent of the action units are randomly activated
- Type 3: All action units are randomly activated.

These single action units were randomly combined with the *Face Randomizer* at a randomly selected activation level between and 0.7. The resulting virtual faces were presented in random order to the participants using a portable computer. In this approach no expression occurred twice and each participant rated different expressions of the same virtual person under the same conditions.

The faces were rated with bipolar adjectives (on a 1 to 7 scale) from the "Semantic Differential Measures of Emotional State or Characteristic (Trait) Emotions" created by MEHRABIAN and RUSSELL (1974) (translated into German), including the dimensions pleasure-aversion, calmness-arousal, and dominance-submission (the latter was dropped for this analysis).

Note that this procedure leads to an immense variance in stimulus presentation—because of the randomization no facial expression was identical to another facial expression. This is the first time where such a variance in facial expression was used in a rating study.

N= 403 male and female students (mean age 22.4) were recruited as participants on a voluntary basis. Each rating situation took approximately 25 minutes; on average each participant rated eleven different faces. Overall N=2904 randomly generated expressive faces were rated. The analysis method for the generation of the expert database which maps single facial actions to arousal and pleasure states has been applied successfully to the analysis of self reported affect in body posture by GRAMMER et.al. (2004). The method uses a radical bottom- up approach by collecting low level data (either body angles for postures or single facial muscle movements) in their context. The independent variable is either an emotional self report or third party ratings. The independent variable then is regressed upon the low level data and then an emotion space can be constructed for the activation of muscles. Thus we can build an activation space for each facial muscle in a pleasure and arousal circumplex model. The model then can be used to translate internal pleasure and arousal states of a simulation into expressive behavior. Basically this method can be used with any behavior.

Regression analysis of arousal and pleasure on AU activation

A factor analysis was conducted on the rating adjectives (Principal components with Varimax rotation) and as expected a three factor solution. The first factor is a pleasure factor (which explains 27 % of variance), the second factor is a dominance factor (24.9 %) and the third factor was the arousal factor (16.7 %). In the further analysis we will restrict ourselves to the pleasure and arousal factors. Finally the regression factor scores for each rated face for pleasure and arousal on the single muscles were calculated. This gives an activation slope for a single muscle for the respective dimensions in the pleasure and arousal circumplex model.

Figure 4: Faces with randomly activated Action units, which were used in the study. (Picture 1 (from left) AU1 0.47, AU9 0.19, AU14 0.56, AU24 0.58, Picture 2 AU5 0.57, AU17 0.33, AU20 0.62, AU23 0.63, Picture 3 AU4 0.3, AU7 0.6, AU10 0.34, AU15 0.19, Picture 4 AU2 0.28, AU4 0.29, AU6 0.66, AU12 0.58, AU15 0.41).

The regression model was calculated for each muscle on the pleasure and arousal factor scores. The two models with 25 regression scores were tested with a Monte-Carlo Simulation against random models, which yielded an error probability below 0.05 for both models. The regression model for each muscle can have a positive and a negative slope. As muscle activation cannot be negative (in real life) we computed two faces for each dimension. All AUs with a positive slope make the positive dimension (arousal and pleasure). All AUs with a negative slope generate the respective opposite dimension (sleepy, not aroused and displeasure).

In faces where either fifty percent or all AUs were randomly activated, arousal is characterized by the following pattern: The presence of AU5, AU4, AU10, AU12, AU16, AU17, AU23 and AU27 indicate arousal. Non-arousal is associated with the presence of AU18 and AU43. In Type 2 and 3 faces the presence of AU2, AU5, AU12, AU13 and AU27 significantly indicate pleasure. Non-pleasure is indicated by AU4, AU6, AU7, AU9, AU10, AU11, AU14, AU15, AU16, AU17, AU20, AU24 and AU43 (See figure 5). The most striking result of this analysis is that facial expressions are not only defined by the occurrence of specific muscle activations, but also by their absence.

In order to generate a simple model for the visualization of the facial expressions, which show either positive or negative values, arousal, or pleasure, we calculated regression lines for every action unit from the values intensity and arousal, and intensity and valence of the actions units. As these calculations assume a linear relation, this model is an approximation to reality. However, we can calculate the expression, which is linked either positively or negatively to pleasure and arousal. (Fig. 5).

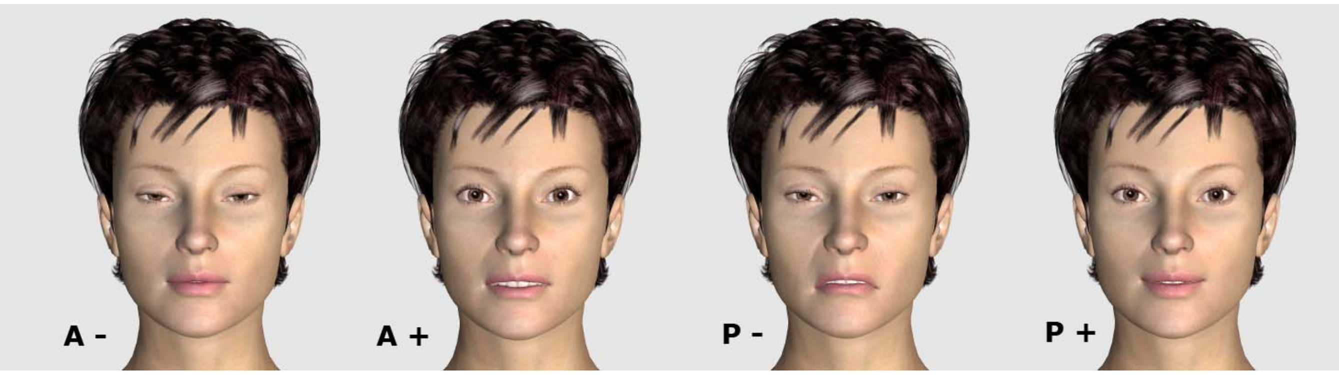

Figure 5: Visualization of the regression analysis of muscle activation upon pleasure (P+), displeasure (P-), arousal (A+) non-arousal (A-). In this case either fifty percent or all AUs were randomly activated. Thus is regression is part of the occurrence of muscle activation in the facial context—i.e. the activation of other muscles.

Two dimensional regression analysis of pleasure and arousal on AU activation

In order to find out if there are distinct pleasure and arousal spaces for each muscle contraction or if contractions are distributed randomly in the pleasure and arousal space a LOESS regression was applied to pleasure and arousal and every AU activation separately (CLEVELAND, 1979; CLEVELAND and DEVLIN 1988).

The single AU regression spaces were all calculated and the regression area was re-sampled (Fig. 6). The data suggest that the pleasure and arousal space is not a linear space for all AUs. In the case of AU1 we see that this AU is activated when pleasure is high and arousal is low, but there is a second peak with medium pleasure and high arousal. This is different for AU5, which is activated under high arousal and high pleasure. AU12 that is responsible for smiling is active under high arousal and high pleasure.

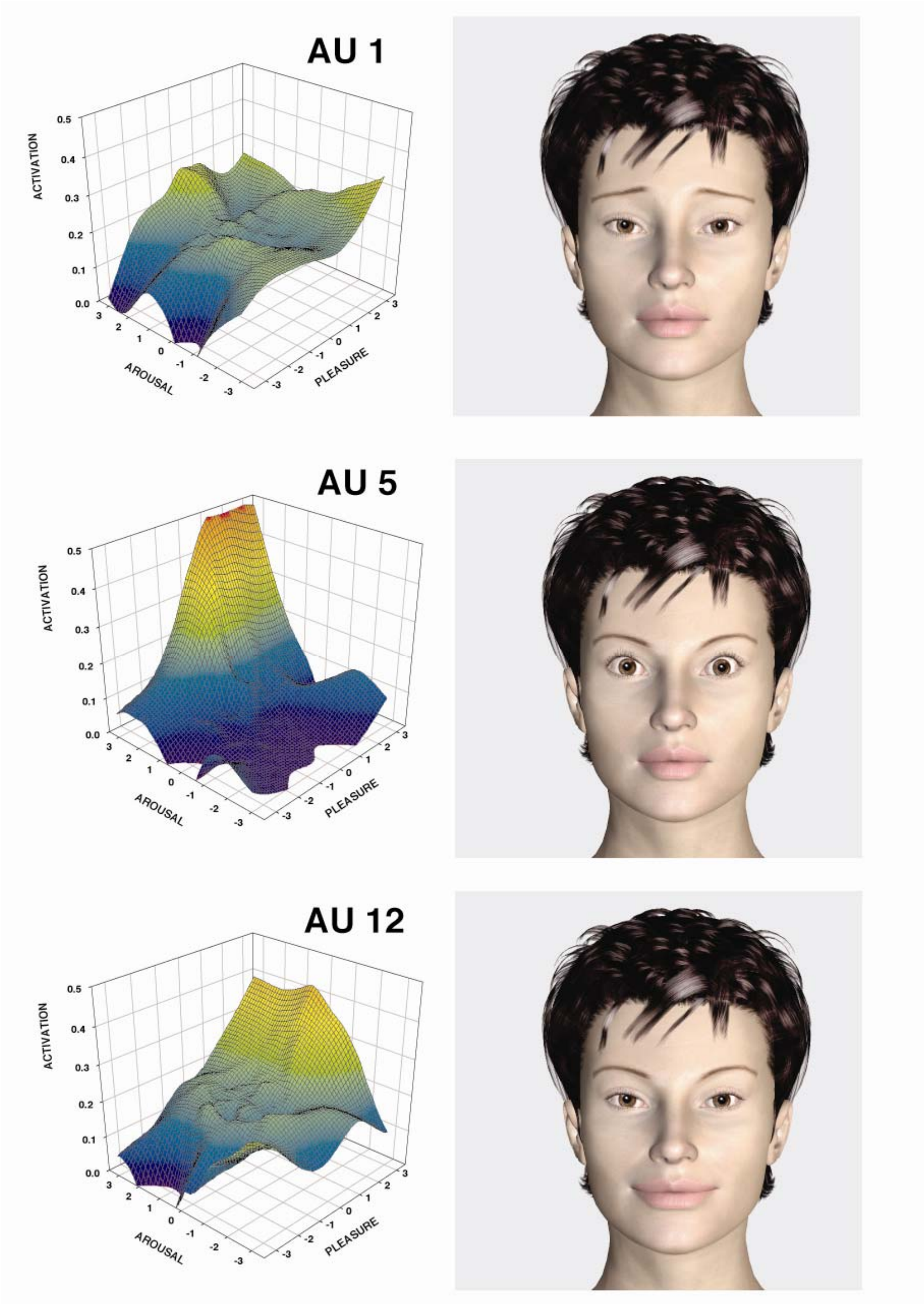

Figure 6: The activation of single AUs in a pleasure and arousal space. Note that these distributions are different for different Action Units.

With this information the complete pleasure and arousal space can be reconstructed by combining all planes for all muscles. The activation values of all AUs at a distinct point in the pleasure-arousal-space can be recomposed into the expression communicating to the respective emotion. Figure 6 shows the results. The pleasure-arousal-spaces of all AUs now can be used to drive a simulation on the basis of only two variables—arousal and pleasure—and create the corresponding facial expression automatically.

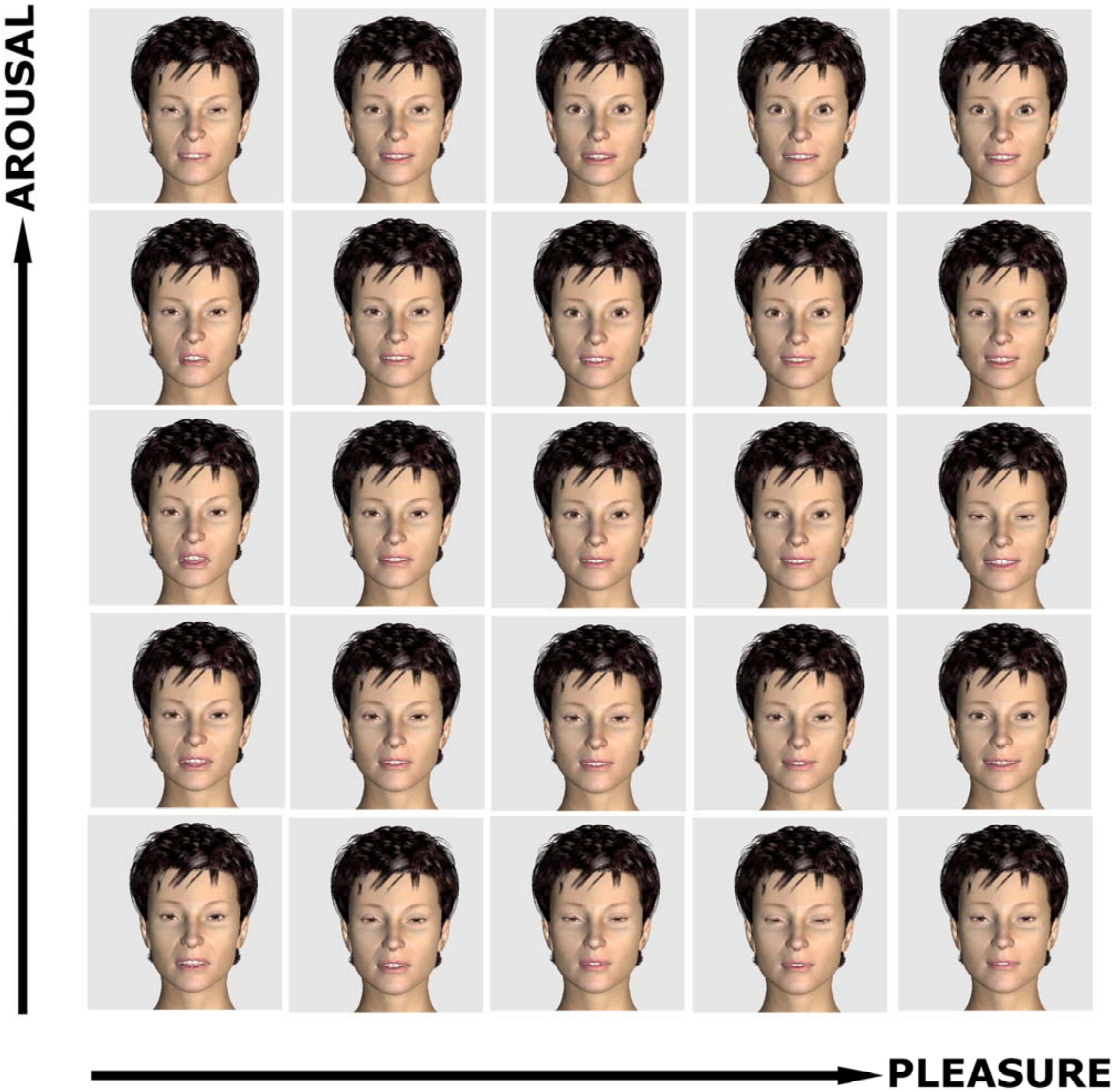

Figure 7: The completely reconstructed pleasure and arousal space.

Reconstruction of basic emotions in the pleasure and arousal space

In a last step we tried to figure out if both models, the *Expression Program* and the *Circumplex Model* actually could be unified at least theoretically. In order to accomplish this we calculated the pleasure and arousal spaces for categorical emotions from the action units, which constitute them. This is a single adding up of all values of activation at any point of the space.

Surprisingly the categorical emotions are not distributed uniformly in the pleasure and arousal space. Surprise (AU1 + AU2 +AU5 +AU26) occurs most often when pleasure is neutral and arousal is high. Sadness (AU1+ AU4 +AU15) coincides with low pleasure and slightly raised arousal, but the plane also indicates that sadness could occur in situation of low arousal and high pleasure. Happiness (AU6 + AU12) occurs in our pleasure and arousal space when pleasure is high and arousal slightly above neutral. Fear (AU1 + AU2 +AU4 +AU20 + AU26) shows relations to low pleasure and high arousal as does disgust (AU9 + AU10). The difference between fear and disgust is that fear does not occur at medium arousal and low pleasure, but disgust does. The last of the six base emotions is anger (AU4 + AU5 + AU7 + AU10 + AU23 + AU26). It occurs under high arousal and low pleasure conditions but also under high arousal and high pleasure conditions and even under low arousal and high pleasure conditions. With these results we are now able to map the categorical model into the *Circumplex Model* and we hypothesize that the first model is a subspace of the latter model.

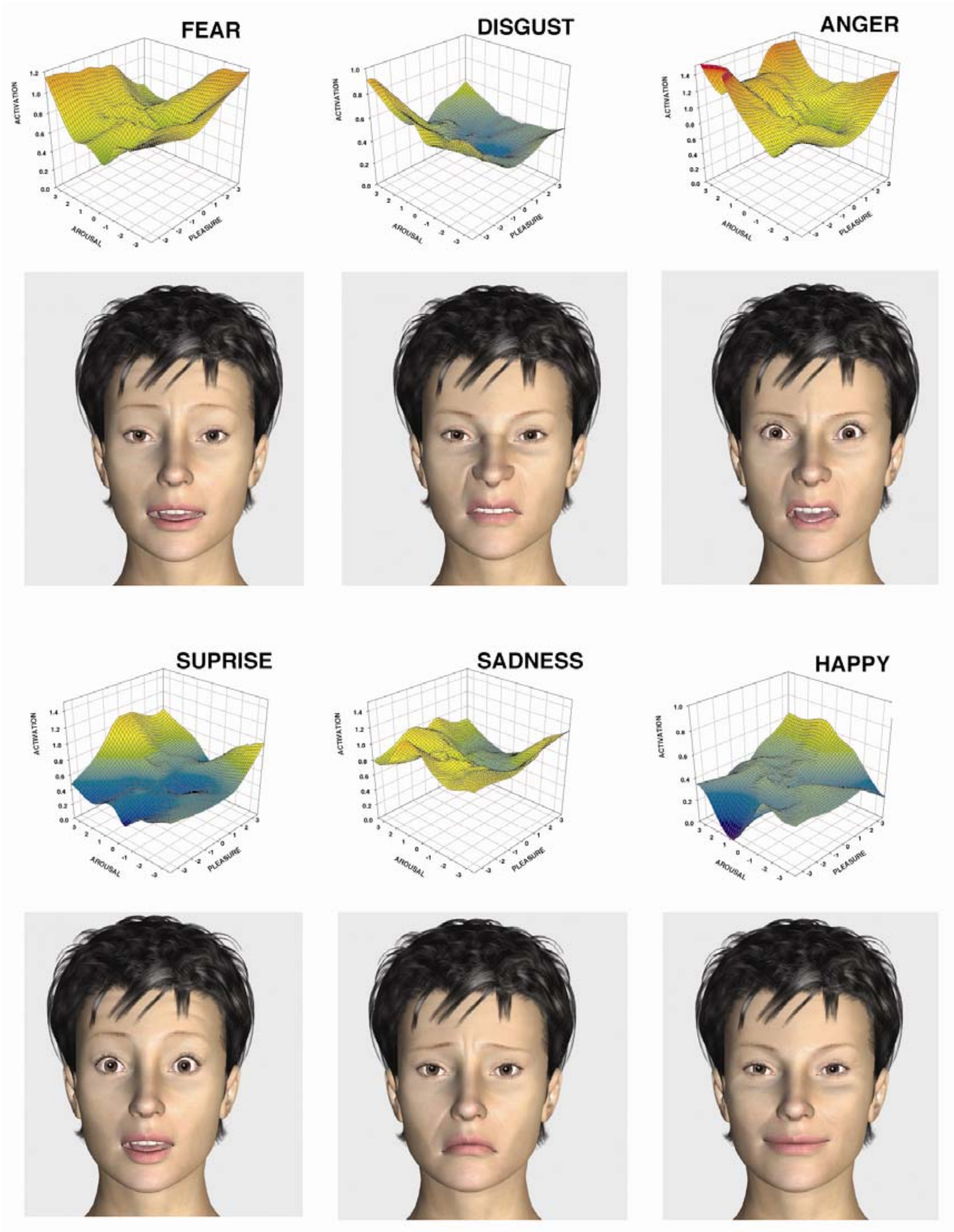

Figure 8: The reconstruction of categorical emotion expressions in the pleasure and arousal space. In comparison to the predictions from figure 1 we find an almost complete match—but there are also exceptions, i.e. the same categorical emotion can occur under different pleasure and arousal combinations.

Facial expression simulation

In this explorative work we have demonstrated that it seems to be possible to create control architecture for expression simulation on the basis of only two controlling variables—pleasure and arousal—which can generate highly complex and variable facial expressions. This architecture can easily be implemented under a variety of conditions. This approach provides a simple solution for the action expression problem in virtual agents and embodied systems.

The system can generate virtually endless combinations of AU activations and thus the expressive behavior of such an agent will never become monotonous. This approach also shows that a radical bottom up approach in nonverbal behavior research is feasible and generates convincing results. In regard to the construction of avatars and embodied systems in our view this is the only approach, which is able to solve the action expression problem.

Besides the finding that the control architecture is feasible we have also demonstrated, that even apparently contradicting emotion theories can be implemented under the same control architecture. For instance it should be possible to weigh action units according to emotional categorical arousal. This means that these action units would be shown preferably, according to the given emotional conditions. With this architecture it would be possible to let an avatar show emotions, which apparently contradict the context—i.e. crying under highly happy circumstances.

Basically all emotion theories can be mapped onto facial expressions with this procedure, as long a semantic differential for this theory can be constructed. It would be plausible for instance to have subjects rate random faces in the six basic emotions dimensions and then calculate the regression values for each AU and base emotions. This demonstrates that our approach is versatile and can be used to test the validity of emotion theories. But there is a general caveat. In the case of random faces there is no information about the internal state of the sender. This information would be necessary to create a complete, accurate and valid model.

Many aspects of expression simulation have not yet been touched by this first approach, for example, the occurrence of facial asymmetries in expression. With our approach this would be easy to implement by simply splitting the morph targets that compose the action units. Control architectures for asymmetries can then be built the same way as we did it for the symmetrical expressions. But actually other semantic differentials can also be used. For example, differentials covering the honesty of the expression would only add one variable to our system.

Another problem not touched at all, even in facial expression research, is the assessment of *dynamics* of action units. But again we suggest that this can be done the same way as for the symmetrical facial expressions. We suggest that it is possible to put all other types of non-verbal behavior—from head and eye movements and gestures, over postures to emotional patterns in speech—in a comparable system. Indeed we have shown this for emotional information in body posture. In this case the angles formed by body joints were subject to the regression with self-described emotional state (GRAMMER et al., 2004).

The *reverse engineering* approach generally would consist of a collection of either simulated or real life data. In the case of real life observation data self-descriptions can be collected parallel. In a second step, observers rate the resulting stimuli and then a regression analysis on the data is made. The prerequisite is that the behavior data are extremely low level, like muscle movements or body angles. Out of these regressions new stimuli on avatars can be created, which can then be rated again and compared to self-description.

However, the present approach has not only consequences for control architecture research but also for the field of non-verbal behavior research itself. If such an approach finds adoption widely, it could help overcoming the stagnation of research in non-verbal behavior. This tool creates the opportunity to address open questions in facial expression research using the new methodological possibilities. Being able to manipulate facial expressions with a high resolution and sufficient realism, the nature and meaning of facial expressions can be investigated in more detail: For example, it is possible to look at the meaning of single AUs more thoroughly—what is the threshold intensity above which they can convey a meaning? By introducing asymmetry, also these phenomena can be studied. Interaction effects of AUs can also be addressed: i.e. which combination of AUs is necessary and / or sufficient to create a convincingly prototypical emotion expression, and what variations in facial expressions are attributed with the same base emotion?

How are facial expressions appraised that are composed of parts of the action units defining a base emotions—this question would discriminate between compulsory and facultative elements of the expressions of base emotions as they are described by EKMAN & FRIESEN (1978).

Using reverse engineering research is no longer limited to expressions of base emotions of rather high intensity. By applying this technique, every combination and intensity of AUs can be investigated. This allows analyzing mixed emotions, i.e. overlaps or combinations of base emotions.

Apart from the broad possibilities in research of static expressions of emotions, reverse engineering opens a new horizon for the time aspect in facial behavior. Timing might be a source of additional information, i.e. not only the resulting expression (as a state), but the special quality of movement leading to this state might convey information about the internal state, as well as motives. In this context, facial expressions can be understood as the apex, which is framed by an onset- and an offset-movement. The relation of the duration of these three periods (onset—apex—offset) might be one source of dynamic information. Above that, the timing of the single AUs within the onset and offset period could modulate the meaning of an expression, i.e. whether they are activated simultaneously or sequentially. Another open question is the existence of conversational signals, which control interactions (FRIEDLUND, 1994). When facial signals occur to emphasize certain parts of speech—the control architecture will become a little more difficult. We suggest that an approach taking the 'Behavioral Ecology View' theory into account simply will add only a new layer to our control architecture, as it is the case for the emotion program theory.

All in all, the present study is only the first step both for research based on reversed engineering and implementing emotional behavior in avatars. Our results hint at the qualities of this approach for both fields: In research on nonverbal behavior the ability to modulate behavioral output (in combination with a metric measurement) creates the opportunity to analyze the qualitative meaning of behavior elements. In behavior simulation, this method allows to implement complex behavior patterns with a comparably low complexity combined with a high plasticity and variability in the output. Thus, both computer engineering and research on the theory of emotions can benefit from this approach.

References

CASSELL, J. (2004): Towards a Model of Technology and Literacy Development: Story Listening Systems. *Journal of Applied Developmental Psychology 25* (1), 75 – 105.

CLEVELAND, W. S. (1979): Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association, Vol. 74*, 829 – 836.

CLEVELAND, W. S.; DEVLIN, S. J. (1988): Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association, Vol. 83*, 596 – 610.

EKMAN, P. (1971): Universals and Cultural Differences in Facial Expressions of Emotion. *Nebraska Symposium on Motivation* 19, 207 – 283.

EKMAN, P. (1984): Expression and the Nature of Emotion. In: K. SCHERER; P. EKMAN (Eds.): *Approaches to Emotion*. Hillsdale, NJ: Lawrence Erlbaum, 319 – 343.

EKMAN, P.; FRIESEN, W. V. (1969): The repertoire of nonverbal behavior—Categories, origins, usage and coding. *Semiotica*, 1, 49 – 98.

EKMAN, P.; FRIESEN, W. V. (1978): *Investigator's Guide: Facial Action Coding System*. Paolo Alto, CA: Consulting Psychologists Press.

FRIDLUND, A. J. (1994): *Human facial expression: An evolutionary view*. San Diego: Academic.

FRIJDA, N. H. (1986): *The Emotions***.** Cambridge: Cambridge University Press.

GILLENSON, M. L. (1974): *The interactive generation of facial images on a CRT using a heuristic strategy*. Ohio State University, Computer Graphics Research Group.

GRAMMER, K.; FINK, B.; OBERZAUCHER, E.; ATZMUELLER, M.; BLANTAR, I.; MITTEROECKER, P. (2004): The representation of self reported affect in body posture and body posture simulation. *Collegium Anthropologicum* (in press).

IZARD, C. E. (1971): *The Face of Emotion*. New York: Appelton-Century-Crofts.

LANG, P. J. (1995): The emotion probe. *American Psychologist, 50*, 372 – 385.

LAZARUS, R. S. (1991): *Emotion and adaption*. Oxford University Press, New York.

MEHRABIAN, A.; RUSSEL, J. A. (1974): *An approach to environmental psychology*. Cambridge: MIT Press.

PANKSEPP, J. (1992): A critical role for 'Affective Neuroscience' in resolving what is basic about basic emotions. *Psychological Review, 99*(3), 554 – 560.

PARKE, F. I. (1972): Computer generated animation of faces. *ACM Nat'l Conference*, 1, 451 – 457.

PETTA P.; STALLER A.; TRAPPL R.; MANTLER S.; SZALAVARI Z.; PSIK T.; GERVAUTZ, M. (1999): Towards Engaging Full-Body Interaction. In: *Proc. of HCI International '99*, Munich, Germany, August 22 – 27.

PLATT, S. M.; BADLER, N. (1981): Animating facial expressio. *Computer Graphics*, *15*(3), 245 – 252.

REEVES, B.; NASS, C. (1996): *The Media Equation. How People Treat Computers, Television, and New Media like real People and Places*. New York, NY: Cambridge University Press.

RUSSELL, J. A. (1978): Evidence of convergent validity on the dimensions of affect. *Journal of Personality and Social Psychology*, 36, 1152 – 1168.

RUSSELL, J. A. (1980) A circumplex model of affect. *Journal of Personality and Social Psychology*, *60*, 37 – 49.

RUSSELL, J. A. (1991): In defense of a prototype apprach to emotion concepts. *Journal of Personality and Social Psychology*, 60, 425 – 438.

RUSSELL, J. A. (1995): Facial expression of emotion: What lies beyond minimal universality? *Psychological Bulletin*, 118, 379 – 391.

SCHERER, K. R. (1984): On the nature and function of emotion: A component process approach. In: K. R. SCHERER; P. EKMAN (Eds.): *Approaches to emotion*. Hillsdale, NJ: Erlbaum, 293 – 317.

SCHERER, K. R. (1999): Appraisal Theory. In: T. DALGLEISH; M. J. POWER (Eds.), *Handbook of Emotion and Cognition*. Chichester: New York, 637 – 663.

SCHERER, K. R. (2001): Appraisal considered as a process of multi-level sequential checking. In: K. R. SCHERER; A.SCHORR; T. JOHNSTONE (Eds.): *Appraisal Processes in Emotion: Theory, Methods, Research*. New York: Oxford University Press, 92 – 120.

SNODGRASS, J. (1992): *Judgment of feeling states from facial behavior: A bottom-up approach*. Unpublished doctoral dissertation, University of British Columbia.

SPENCER-SMITH, J.; WILD, H.; INNES-KER, A. H.; TOWNSEND, J. T.; DUFFY, C.; EDWARDS, C.; ERVIN, K.; MERRITT, N.; PAIK, J. W. (2001): Making faces: Creating three-dimensional parameterized models of facial expression. *Behavior Research Methods, Instruments and Computers*, 33, 115 – 123.

STALLER A.; PETTA P. (1999): Towards a Tractable Appraisal-Based Architecture for Situated Cognizers. In: CANAMERO, D. et al. (Eds.) *Grounding Emotions in Adaptive Systems, Workshop Notes, Fifth International Conference of the Society for Adaptive Behaviour (SAB98)*, Zurich, Switzerland, August 21, 1999.

TOMKINS, S. S. (1962): *Affect imagery consciousness: Vol. 1. The positive affects*. New York: Springer.

WEHRLE, T.; SCHERER, K. R. (2001): Towards computational modeling of appraisal theories. In: K. R. SCHERER; A. SCHORR; T. JOHNSTONE (Eds.): *Appraisal Processes in Emotion: Theory, Methods, Research*. New York: Oxford University Press, 350 – 365.