Are computer-generated emotions and moods plausible to humans?

Patrick Gebhard[1], Kerstin H. Kipp[2]

[1]German Research Center for Artificial Intelligence,
[2]Experimental Neuropsychology Unit, Saarland University,
Saarbrücken, Germany
[1]patrick.gebhard@dfki.de, [2]k.kipp@mx.uni-saarland.de

Abstract. This paper presents results of the plausibility evaluation of computer-generated emotions and moods. They are generated by ALMA (A Layered Model of Affect), a real-time computational model of affect, designed to serve as a modular extension for virtual humans. By a unique integration of psychological models of affect, it provides three major affect types: emotions, moods and personality that cover short, medium, and long term affect. The evaluation of computer-generated affect is based on textual dialog situations in which at least two characters are interacting with each other. In this setup, elicited emotions or the change of mood are defined as consequences of dialog contributions from the involved characters. The results indicate that ALMA provides authentic believable emotions and moods. They can be used for modules that control cognitive processes and physical behavior of virtual humans in order to improve their lifelikeness and their believable qualities.

1 Introduction and Related Work

The employment of virtual humans as an interface in a human-computer interaction scenarios makes high demands on their believability. This is mainly founded in the sophisticated embodiment of virtual humans that implies human-like conversational abilities and behavior aspects. In general, they can be defined as the embodiment of all internal processes through non-verbal and verbal aspects. Such are the character's gestures, posture, speech characteristics, and facial complexions. On a higher level these internal processes comprise the character's cognitive processes like decision making, subjective appraisal of situations, and dialog strategy selection that are responsible for the conversational behavior in an interaction. In order to increase the believability of a face-to-face virtual conversation partner, researchers have begun to address the modeling and emulation of human-like qualities such as personality and affective behavior. Examples of such systems are COSMO [1], Émile [2], Peedy [3], and the Greta agent [4]. In general, the generation of affective behavior relies on the OCC emotion model that has been defined by Ortony, Clore, and Collins [5]. As a next step, a few research groups have started to address emotion modeling in multi-character scenarios. In the context of a military mission rehearsal application Traum and Rickel [6] address dialog management comprising human-character and character-character dialogs in immersive virtual environments. Prendinger et. al. [7] developed a framework for scripting presentations with multiple affective characters in a web-based environment. Part of their work is the SCREAM system that computes affective states based on the OCC-model but also considers aspects of the social context, i.e. role and status of characters.

Another enhancement addresses the emotion based generation of behavior and conversational abilities of virtual characters. Therefore, the used model of emotions has been extended by other affective characteristics that can be exploited for the generation of believable behavior. An important extension consists of the modeling of mood, which represents a diffuse, longer lasting state of affect. Lisetti and Gmytrasiewicz [8] focus on the social expertise of agents. They use a hierarchical model of affect, that consists of emotions, mood, and personality for computing a dynamic emotional state by using probabilistic Markov Models. This should enable the design of autonomous, socially intelligent agents that can predict the emotional state of others. However, they focus on the modeling of emotions that control an agent's actions. The approach from Mehdi et al. [9] uses a one dimensional mood space (good vs. bad) as an affective filter that regulates the intensity of the actual emotion. The Artificial Emotion Engine from Wilson [10] was developed to enhance virtual characters in games and other virtual environments. A one dimensional (good vs. bad) mood model that is driven by punishment or reward signals is used to model a virtual characters motivations. For modeling behavior of synthetic actors, Rousseau and Hayes-Roth [11] rely on a social-psychological model that defines the personality of synthetic actors by moods and attitudes. Moods consist of emotions, triggered by external events and sensations that are event-independent, divided in self-oriented moods and agent-oriented moods. In terms of the OCC model, these moods are rather emotions that represent either the appraisal one's own situation or another one's situation.

Following the motivations of other researchers to enhance conversational abilities and behavior aspects of our virtual characters, our group has started to incorporate personality and affective states to extend a character's conversational and social repertoire. We use affective states to color simulated dialogs through verbal and non-verbal expression of emotions [12]. Focusing on multi-party conversations (rather than performing physical actions), emotions can be used in the dialog generation processes to inform the selection of dialog strategies and linguistic style strategies as proposed by [13]. They also play an important role in the turn-taking behavior (e.g. a spontaneous barge-in may result from an intensive emotion) and in the realization of concrete dialog moves. At some point in our research, we realized that relying solely on emotions might not be sufficient to control behavior aspects like gestures, speech characteristics, or communicative abilities. Inspired by Davidson's thesis that „emotions bias action, whereas moods bias cognition“ [14] we enhanced our OCC based computational model of affect by adding „longer-living“ moods to „short-living“ emotions and personality [15]. We are convinced that mood is a complex affect type, like emotions are. Mood modeled as a one-dimensional value will therefore not be appropriate for a rich exploitation in cognitive processes.

From a technical perspective, we do not rely on a full-fledged affective reasoning process to infer emotions, like the affective reasoner by Elliot [16] or the domain-independent framework for modeling emotion by Gratch and Marsella [17]. Our approach of the generation of affect relies on so-called *appraisal tags* [12, 15] that are used to appraise situational events (e.g. a notification about a prize in a lottery), actions (e.g. someone rescues a child from drowning), objects (e.g. a beautiful butterfly), and on a more abstract level to appraise dialog acts (e.g. to tease, or to praise someone) and affect clues (e.g. blush of shame). In general, appraisal tags can be considered as the final output of a higher-level appraisal reasoning process.

In our opinion, the use of appraisal tags facilitates the generation of affect in both, script-controlled and plan-based virtual character systems. An example for of a script-controlled application using appraisal tags for affect generation is CrossTalk [12]. The VirtualHuman application [15,18] is an example for a more sophisticated system of autonomous virtual humans that use a plan-based higher-level appraisal process to generate appraisal tags (as an intermediate stage) to compute emotions and moods that again influence the virtual humans' behavior.

Before affect-based behavior of virtual characters can be checked for plausibility, the underlying models of the behavior generation process should produce meaningful results. At an abstract-level, a virtual character's behavior should be consistent and believable according to the current situation. First, it must be verified that the generated affect is plausible to the current situation and second that the behavior is consistent with respect to the generated affect. This two step behavior generation is illustrated by figure 1. The relation *has to be consistent with (behavior, situation)* implies for the affect-based behavior generation that affect *has to be consistent with* the situation and behavior *has to be consistent with* affect. In our opinion, this has to be a basic step before aspects of a virtual character's behavior can be evaluated according to common evaluation methods collected, for example, by Ruttkay and Pelachaud [19].

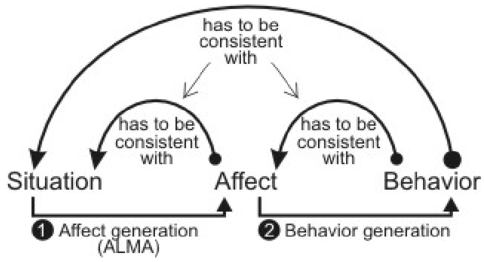

Fig. 1. Consistency relations of affect-based behavior generation.

2 A Layered Model of Affect

One of the challenges in creating a computational model for the mentioned affect types were the modeling of mood, the „longer-living“ affect type. We felt that mood modeled as a one-dimensional value (i.e. good mood vs. bad mood) will not be appropriate for a rich exploitation in cognitive processes. Many interesting behavior aspects might occur, when characters are in a mood out of the good-bad mood dichotomy. Consider, for example, the implications on cognitive processes of the moods anxious or bored. All these requirements raise the question for a model of mood that covers most (or at best all) moods that occur in human beings and that defines how moods can be changed. As a first answer, we have introduced a layered model of affect, which we call ALMA [15] that is based on different psychological models. It comprises three interacting kinds of affect as they occur in human beings:

1. *Emotions* reflect short-term affect. Emotions usually decay and disappear of the individual's focus [18].
2. *Moods* reflect medium-term affect. Moods are longer lasting stable affective states, which have a great influence on human's cognitive functions [21].
3. *Personality* reflects individual differences in mental characteristics (long-term affect). Those can be described by the Big Five model of personality with its traits openness, conscientiousness, extraversion, agreeableness and neuroticism [22].

ALMA is an extension to the computational model of emotions of the EmotionEngine [12,13]. It implements the OCC model of emotions [5] combined with the Big Five factor model of personality [22]. OCC is a cognitive model of emotions, and is essentially based on the concepts of appraisal and intensity. The individual is said to make a cognitive appraisal of the current state of the world. Emotions are defined as valenced reactions to events of concern to an individual, actions of those s/he considers responsible for such actions, and objects/persons. ALMA extends the EmotionEngine by a computational model of mood based on Mehrabian's mood theory [23]. It defines mood as „an average of a person's emotional states across a representative variety of life situations".

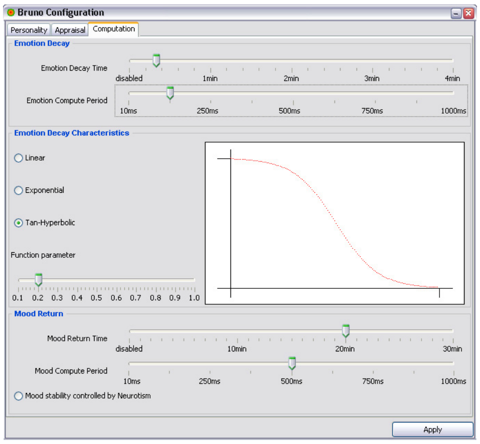

Fig. 2. ALMA affect computation configuration interface.

Initially, personality values are used for the computation of the initial emotion intensities. The current ALMA version uses mood values to increase or decrease the intensity of elicited emotions in order to realize a more natural emotion intensity computation. For example, individuals experience stronger joy emotions when being in an exuberant mood, as when being in a hostile or anxious mood (see also section Mood Change below). In general, the intensity of emotions underlies a natural decay, which is configurable by several decay functions (linear, exponential, and tan-hyperbolical). A graphical interface, see figure 2, provides extensive control about almost all parameters that impacts the affect computation.

Mood Model

Mehrabian describes mood with the three traits *pleasure* (P), *arousal* (A), and *dominance* (D). The three nearly independent traits form a 3-dimensional mood space. The implementation of the PAD mood space uses values from -1.0 to 1.0 for each dimension. Mood is described with the following classification for each of the three mood space axes: +P and –P for pleasant and unpleasant, +A and –A for aroused and unaroused, and +D and –D for dominant and submissive. With this classification all mood octants of the PAD mood space are described by Table 1.

While a point in the PDA space represents the mood, a mood octant represents the discrete mood of an individual. For example, a person's (discrete) mood is relaxed, if the value of P is positive, the value of A is negative, and the value of D is positive. We define the strength of a current mood by its distance to the PAD zero point. The maximum distance is √3. This is divided into 3 equidistant sections that describe three discrete mood intensities: slightly, moderate, and fully. Before mood changes can be computed, it is essential to define an individual's *default mood*. The mapping presented in [24] defines a relationship between the big five personality traits and the PAD space. Relying on this mapping, the EmotionEngine, which uses the big five personality model to define a characters personality, is thereby able to derive a default mood for characters:

Table 1. Mood octants of the PAD space.

+P+A+D	**Exuberant**	-P-A-D	**Bored**
+P+A-D	**Dependent**	-P-A+D	**Disdainful**
+P-A+D	**Relaxed**	-P+A-D	**Anxious**
+P-A-D	**Docile**	-P+A+D	**Hostile**

P := 0.21•Extraversion + 0.59•Agreeableness + 0.19•Neuroticism
A := 0.15•Openness + 0.30•Agreeableness - 0.57•Neuroticism
D := 0.25•Openness + 0.17•Conscientiousness + 0.60•Extraversion - 0.32•Agreeable.

For example, an individual whose personality is defined with the big five personality traits openness=0.4, conscientiousness=0.8, extraversion=0.6, agreeableness=0.3, and neuroticism=0.4 has the default mood slightly relaxed (P=0.38, A=-0.08, D=0.50).

Mood Change

A more challenging task is the simulation of human-like mood changes. Morris [21] has identified four factors that play a role in human mood. All of them are closely related to an emotional experience. To keep the modeling of mood changes as lean as possible, we use emotions as mood changing factors. In order to realize this, emotions must be somehow related to a character's mood. We rely on Mehrabian's mapping of emotions into the PAD space [24]. However, not all 24 emotion types provided by the EmotionEngine are covered by this mapping. For those that lack a mapping, we provide the missing pleasure, arousal, and dominance values by exploiting similarities to comparable emotion types (see Table 2).

For the human-like imitation of mood we rely on a functional approach. We concentrate on how emotions influence the change of the current mood and we consider the aspect that a person's mood gets the more intense the more experiences the person makes that are supporting this mood. For example, if a person's mood

Table 2. Mapping of OCC emotions into PAD space.

Emotion	P	A	D	Mood Octant
Admiration	0.4	0.3	-0.24	+P+A-D Dependent
Anger	-0.51	0.59	0.25	-P+A+D Hostile
Disliking	-0.4	-0.2	0.1	-P-A+D Disdainful
Disappointment	-0.3	-0.4	-0.4	-P-A-D Bored
Distress	-0.4	0.2	0.5	-P+A+D Hostile
Fear	-0.64	0.60	0.43	-P+A-D Hostile
FearsConfirmed	-0.5	0.3	-0.7	-P-A-D Anxious
Gloating	0.3	-0.3	-0.1	+P-A-D Docile
Gratification	0.6	-0.3	0.4	+P-A+D Relaxed
Gratitude	0.2	0.5	-0.3	+P+A-D Dependent
HappyFor	0.4	-0.2	-0.2	+P-A-D Docile
Hate	-0.4	-0.2	0.4	-P-A+D Disdainful
Hope	0.2	0.2	-0.1	+P+A-D Dependent
Joy	0.4	0.2	0.1	+P+A+D Exuberant
Liking	0.40	-0.16	-0.24	+P-A-D Docile
Love	0.3	0.1	0.2	+P+A+D Exuberant
Pity	-0.4	-0.2	-0.5	-P-A-D Bored
Pride	0.4	0.3	0.3	+P+A+D Exuberant
Relief	0.2	-0.3	-0.4	+P-A-D Docile
Remorse	-0.3	0.1	-0.6	-P+A-D Anxious
Reproach	-0.3	-0.1	0.4	-P-A+D Disdainful
Resentment	-0.2	-0.3	-0.2	-P-A-D Bored
Satisfaction	0.3	-0.2	0.4	+P-A+D Relaxed
Shame	-0.3	0.1	-0.6	-P+A-D Anxious

can be described as slightly anxious and several events let the person experience the emotion fear, the person's mood might change to moderate or fully anxious.

Our computation of mood changes is based on active emotions generated by the EmotionEngine. Each appraisal of an action, event or object, lets the EmotionEngine generate an active emotion that once generated, is decayed over a short amount of time (i.e. one minute). All active emotions are input for the mood function. The function has two scopes. Based on all currently active emotions the function defines whether the current mood is intensified or changed. It will be intensified if all active emotions are mapped into the mood octant of the current mood. A mood will be changed progressively if all active emotions are mapped into a different mood octant than the current mood. The mood function is visualized by ALMA within the AffectMonitor that is shown in figure 3. It shows the situation, where the current mood will be changed by active emotions. A detailed description of the mood function can be found in [15].

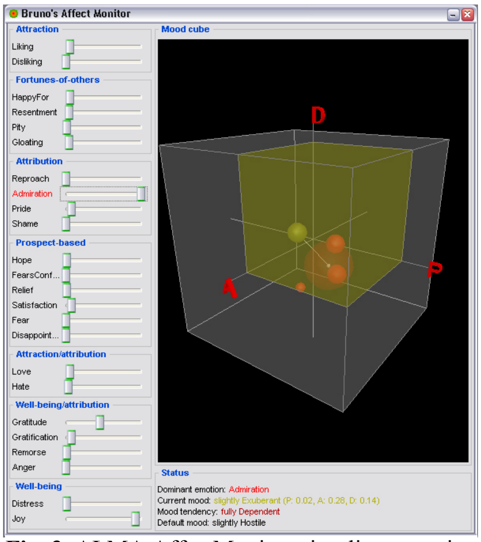

Fig. 3. ALMA AffectMonitor visualizes ongoing mood changes and elicited emotions.

A novelty of the actual version of ALMA is that the current mood influences the intensity of active emotions. The theory is that the current mood is related to personality values that interfere with a character's actual personality values. Technically, this is realized by the reverse use of the (above shown) mapping of big-five personality values on PAD values. Based on the current mood, its temporary virtual personality values will increase or decrease a characters personality values. Those will be used to regulate the intensity of emotions. This increases, for example, the intensity of joy and decreases the intensity of distress, when a character is in an exuberant mood.

Another mentionable aspect is that the current mood has a tendency to slowly move back to the default mood. Generally, the return time depends on how far the current mood is away from the default mood. We take the longest distance of a mood octant ($\sqrt{3}$) for defining the mood return time. Currently this is 20 minutes.

Affect Computation

For the affect computation, ALMA provides a rule based appraisal mechanism. A set of appraisal rules will map affect input to internal values that are used to compute emotions. So called *appraisal tags*, cf. [13], are the symbolical representations of affect input and will be processed by ALMA. We distinguish three types of affect input: 1) *basic appraisal tags*, 2) *act appraisal tags*, and 3) *affect display appraisal tags*. Basic appraisal tags express how a speaking character appraises the event, action or object about which it is talking. Act appraisal tags describe the underlying communicative intent of an utterance, e.g. tease, or congratulate. Affect display appraisal tags are visual cues of an experienced emotion or mood, e.g. a blush of shame or a character that looks nervous for a specific amount of time. Consider the example in which the utterance from Anne is tagged by a human annotator with the act appraisal tag `[Admire Anne]`, which indicates that Anne admires Bruno for something:

Anne: *Bruno, you are dancing like a god. I could dance with you for hours* `[Admire Bruno]`

Each involved character (Bruno, Anne) appraise the appraisal tag by its own set of appraisal rules. The appraisal tag `[Admire Bruno]` is appraised by Anne as `GoodActOther`, a praiseworthy action of another person. Bruno appraises it as `GoodActSelf`, a praiseworthy action of him self. `GoodActOther` and `GoodActSelf` are basic appraisal tags that will resolve into OCC variables for computing emotions. Following the OCC emotion theory a praiseworthy action of oneself will elicit the emotion pride and a praiseworthy action of another person will elicit the emotion admiration. The intensity of the emotions depends on the characters personality values and current mood. For a detailed overview, how we generate emotions by appraisal tags, see [13]. The elicited emotions will impact the mood of each character, according to the mood change function described above. However, more than one (intense) emotion is needed to change a character's current mood.

3 Evaluation

In order to prove that ALMA's computational model of affect is able to produce coherent affect that is comparable to human affect, we ask people how plausible they perceive the generated emotions and moods. To eliminate all (or at least most of the) side-effects that might blur the results, we decided not to evaluate the plausibility of affect through the visualization by a virtual character. The visualization of a specific emotion or mood might be recognized differently from the affect that ALMA has generated. Therefore, we rely on a textual description for generated emotions and moods in the plausibility test. If we could show at this level that the affect generated by ALMA seems plausible, the visualization of them – if done correctly – will be plausible as well.

Methods

We check the plausibility of affect with an offline textual questionnaire that is organized in two sections: one for emotions and one for moods. In the two sections we let the participants judge the plausibility of 24 emotion types and 8 different moods. These are all the affect types that ALMA is able to generate. The sections hold dialog contributions from which we claim that they have an impact on the affect of the involved speaker and addressee.

Participants

In order to investigate relationships regarding affect and participants, we rely on 33 people at different age and gender. The youngest participant was 18 years and the oldest participant was 38 years old. Basically the participants can be divided into two groups: the student group consisting of 17 people between the age of 18 and 19, and the adult group consisting of 16 people between the ages of 25 to 38. Both groups consist of half men and half women.

Material

The materials we use for the evaluation consist of single *dialog contributions*, and *dialog scenes* that can be defined as a set of dialog contributions. Single dialog contributions and dialog scenes have to be annotated with appraisal tags (see section affect computation). Those tags are not shown in the final questionnaire, but used to compute the affect related to the dialog contribution or dialog scene. In addition, we rely on a set of appraisal rules for each involved character, which is needed by ALMA to compute affect.

The basic assumption we made is that emotions will be elicited by dialog contributions. For example, the dialog contribution of Bruno *„Anne, I know you are well prepared for your exam. I am sure you will pass it with a good grade*“ elicits the emotion hope on the side of Anne, the addressee. On the side of the speaker (Bruno) the emotion pride is elicited. In general, the elicited emotion type depends on how an individual appraises the dialog contributions. The relationship between characters has a great impact how they appraise actions of each other. We define that all characters like each other, but all dislike the character Clementine. This information is explicitly mentioned in the final questionnaire.

Therefore, they have to be enriched by appraisal tags, which stand for the intentional content (see section Affect Computation). These appraisal tags can be processed by ALMA for generating emotions.

Taking the example above, in which Bruno encourages Anne for her exam, the enriched version of the dialog contribution looks like:

Bruno: *Anne, I know you are well prepared for your exam. I am sure you will pass it with a good grade.* `[Encourage Anne]`

Only the act tag is used as input for ALMA. As described above, each character has a set of appraisal rules (about 30-50), which appraise the act tag by taking into account the role of the individual. The act tag `[Encourage Anne]` is appraised by Bruno as `GoodActSelf`, a praiseworthy action of himself, whereas Anne appraises the act tag as `GoodLikelyFutureEvent`, a desirable likely future event that Bruno has put into her mind by saying the above line. Following the OCC emotion theory a praiseworthy action of oneself will elicit the emotion pride (Bruno) and a desirable likely future event will elicit the emotion hope (Anne).

According to Morris' theory (see section Mood Change), which is implemented by ALMA, emotions influence the current mood. Emotions elicited by a set of dialog contributions in a specific time interval can change the current mood of an individual to another mood. For the questionnaire, we use short (mostly singular) dialog contributions for the elicitation of emotions and dialog scenes for the change of moods. For the plausibility check of the 24 emotion types, we rely on 24 short dialog contributions that influence both, the speaker's and the addressee's emotions, see figure 4. Therefore, on average, each emotion is rated 2 times. For the plausibility check of the 8 mood types, we rely on 24 dialog situations, see figure 5. Thus, every mood type is rated 3 times.

Bruno: Anne, it's cool that you're helping grand-mother in cleaning up the garden!

Anne's emotion: *pride* **Bruno's emotion**: *admiration*

Fig. 4. Dialog contributions for emotions.

Situation: Mark is reorganizing his computer hard drive by letting Microsoft Windows removing unneeded files. Tanja just shows up.

Mark: Crap, Windows has killed all pictures of our last summer holiday at Mallorca.
Tanja: Don't panic, you'll find them surely in the waste bin.
Mark: Are you sure? But what if not, what I'm doing then – they will be lost forever!
Tanja: Well, I've no clue, I'm not the computer expert.
(Mark tries to recover the files by restoring the files of the waste bin)
Mark: No, damn it! All the pictures gone – and there's no way to get them back!
Tanja: Oh no, All our pictures are lost! You are a clean up maniac. I always told you that this will led some days to something bad. Well, and that's just happened. Wonderful!
Mark: Get of my back!

Marks mood after: *hostile*

Fig. 5. Dialog scenes for moods.

Because the questionnaire is made for native-speakers, all dialog contributions and dialog situations are originally written in German.

ALMA is used to compute emotions and mood changes based on annotated dialog contributions or dialog situations. The annotation process is part of the evaluation procedure and will be described in the next section. Due to the fact, that ALMA is designed to compute real-time affect comparable to humans, we have to take the duration of a dialog contribution into account. Basically this is only relevant for the dialog scenes. A timer, which is waiting some seconds considering the length of the current contribution, will activate the next dialog contribution as input for ALMA. In the questionnaire, the computed emotions for each individual are noted below a dialog contribution (see figure 4). In case of a dialog situation, the changed mood of the main character is noted below the situation (see figure 5).

Procedure

In a pre evaluation, experts (a computer linguist, a dialog expert, and a psychologist) have reviewed the dialog contributions and the dialog situations for being realistic. All problematic formulations, unrealistic contributions, and unclear situations have been rewritten and modified.

Table 3. Description of moods.

Mood	Description
Exuberant	Extroverted, outgoing, happy, sociable
Bored	Sad, lonely, socially withdrawn, physically inactive
Relaxed	Comfortable, secure, confident, resilient to stress
Anxious	Worried, nervous, insecure, tense, unhappy, illness prone
Dependent	Attached to people, needy of others and their help, interpersonally positive and sociable
Disdainful	Contemptuous of others, loner, withdrawn and calculating, sometimes anti-social
Docile	Pleasant, unemotional, and submissive; likeable; conforming
Hostile	Angry, emotional in negative ways, possibly violent

In a next step, the annotated appraisal tags that represent the intentional content are reviewed for being appropriate. All inappropriate tags have been identified and changed. Based on this material emotions and moods are computed by ALMA. The resulting affect type is displayed below the dialog contributions for the respective dialog situations.

Participants are asked to evaluate during a period about half an hour how plausible emotions and moods are through a discrete ranking scale. The ranking scale was explicitly explained. In the final questionnaire, no more information is given, apart from the list of all possible emotions and moods. Only moods are explained by attributes (translated in German) according to Mehrabian's description of moods (see Table 3).

Data Analysis

In the questionnaire all rankings consisted of a discrete 1-5 scale, 1 denotes the „lowest plausibility", and 5 stands for the „highest plausibility".

Since rating scales can be treated as interval scales [25], we used parametrical tests for the statistical analysis. The *t-test for one sample* is a statistical significance test that proves whether a measured mean value of an observed group differs from an expected value. In our study, ratings were proven to be „positive" if the mean score significantly exceeded the moderate plausible value of 3.

To test the effect of a factor with multiple values (e.g. emotion type) or interactive effects of several factors (e.g. affect type and gender) we calculated *an analysis of variance* (*ANOVA*).

Results

Besides the plausibility score and the plausibility significance of computer-generated emotions and moods, we want to know if age or gender has an impact on the rating.

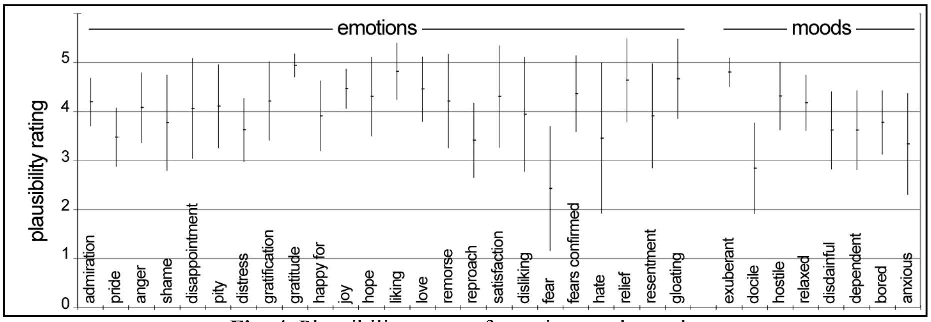

Fig. 4. Plausibility score of emotions and moods.

Figure 4 shows the mean plausibility score and the related standard deviation for each emotion and mood. The following subsections present the details of our plausibility evaluation.

Differences in plausibility rating due to age

To check if differences in the plausibility rating are somehow related to the participants' age a 2 x 2 ANOVA with the factors age group (students, adults) and affect type (emotion, mood) was calculated. Only the factor affect type revealed significance ($F(1,31)=23.38$; $p<.001$) which demonstrates that modeled emotions (Mean=4.07) were rated significantly more plausible than modeled moods (Mean=3.81). Neither the factor age ($F(1,31)=1.67$; $p=.21$) nor the interaction between affect type and age ($F(1,31)=1.57$; $p=.22$) showed a significant effect. Thus, the two age groups did not differ in respect of the rating of emotions and moods and the two age groups were merged for further analysis.

Differences in plausibility rating due to gender

A 2 x 2 ANOVA with the factors gender (male, female) and affect type (emotion, mood) only revealed the above mentioned significant main effect for affect type ($F(1,31)=21.89$; $p<.001$). There was no effect of the factor gender ($F(1,31)=.79$; $p=.38$) and no interaction between affect type and gender ($F(1,31)=.01$; $p=.93$). This reveals that male and female participants did not rate the plausibility of emotions and moods differently. Therefore, both groups were merged for further analysis.

Emotions and Moods

Analyzing different types of emotions in a one-way ANOVA revealed a main effect of the factor „type of emotion“ with 24 levels ($F(23,736)=14.29$; $p<.001$). This demonstrates that the different emotions were rated differently plausible with maximum score for gratification (Mean=4.78) and a minimum score for fear (Mean=2,42).

An equal one-way ANOVA for the factor „type of mood“ with 8 levels shows a significant main effect ($F(7,224)=26.63$; $p<.001$). That is, also modeled moods were rated to be differently plausible. The best score was achieved for exuberant (Mean=4.78) and the worst for dependent (Mean=2.84).

Plausibility of emotions and moods

Comparing the mean of the plausibility ratings of all emotions with the neutral score of 3 demonstrates that the modeled emotions are scored significantly above median ($t(32)=18.63$, $p<.001$).

The analysis of single emotions shows that except for the two emotions hate and fear all emotions are scored positive (for all: $p<.001$). The score for fear was significantly lower than 3 (Mean=2.42; $t(32)=-2.59$; $p<.05$). Detailed analysis showed that adults scored the plausibility of the modelled fear neutrally (Mean=3.00). However, students' score was negative (Mean=1.81; $t(15)=-4.28$; $p<.001$).

The score for hate did not differ from the neutral score of 3 (Mean=3.45; $t(32)=1.69$; $p=.10$). This is due to the fact that the adult group's rating was neutral (Mean=3.12; $t(16)=.32$; $p=.76$) whereas the student group's rating was clearly positive (Mean=3.81, $t(15)=2.14$; $p<.05$).

The mean rating score averaged over all mood types was significantly higher than the neutral score 3 ($t(32)=11.32$, $p<.001$). Thus, the plausibility of moods was generally rated positive.

The analysis of single moods revealed positive scores for all mood types (for all: $p<.001$) except for dependent and anxious. The mood dependent was rated mean (Mean=2.84; $t(32)=-1.00$; $p=.33$). The mood anxious shows a tendency to a positive rating (Mean=3.33; $t(32)=1.84$; $p=.08$). A separate examination for both age groups demonstrates that adults rate the plausibility positive (Mean=3.88; $t(16)=5.08$; $p<.001$), whereas students give a mean rating (Mean=2.75; $t(15)=.97$; $p=.35$).

Discussion
A first observation of the results reveals that emotion types are perceived more significantly plausible than mood types in general. We relate this to the fact that emotion types are used for the computation of mood types. The „plausibility weakness“ of emotion types will be „transferred“ on related mood types according to the mood change functions of ALMA. However, there is no obvious functional description between the plausibility of emotions and related mood.

The bad performance of the emotions fear and hate, which are below neutral and neutral, lacks a well-founded explanation. In case of the emotion fear which is rated better by the adult group, our ad-hoc interpretation was that students have another perception of fear than adults. The related dialog contribution contains a statement in which the addressee's job is about to be recalled if the addressee is not doing its work correctly. Maybe this topic is not (yet) very significant for a student and it is therefore not plausible for them that the addressee experiences fear.

The plausibility rating for the mood dependent was mean. According to our first impression, this might be caused by the German translation of the mood word. In an informal interview this was approved by participants. However, this could also be an effect that is based on the wrong annotation of the dialog contributions of the respective dialog scenes or on the wrong correlation of the elicited emotions to the mood dependent.

6 Conclusion

In this paper we have presented an evaluation of the plausibility of generated emotions and moods by the extended version of ALMA. This version provides a more natural emotion intensity computation because it considers both the personality of an individual as well as its current mood.

We have done a plausibility evaluation of all 24 emotions and 8 moods that can be generated by ALMA. In order not to blur the results, we have explicitly not used virtual characters to visualize emotions and moods. Consequently, the evaluation participants have been confronted with a textual description of the generated emotions and moods.

The overall result of the evaluation is that emotions and moods generated by ALMA are plausible. Considering all participants, the results are independent from age or gender.

Based on these results, we are expecting that the embodiment of ALMA generated emotions and moods through virtual characters will be plausible as well, and will serve well for the generation of behavioral aspects of virtual characters. In addition, the appraisal tag interface allows an easy integration of ALMA's affect computation in script or plan-based virtual character applications.

Acknowledgements

For being a great help in every situation, I would like to thank my colleagues Martin Klesen and Michael Kipp. Also, I would like to thank my brother Gernot Gebhard for coding parts of the ALMA software. The present work is funded by the EU as part of the IST 6th framework network of excellence HUMAINE (Contract no. 507422), and the VirtualHuman project funded by the German Ministry for Education and Research under grants 01 IMB 01A.

References

[1] Lester J., Voerman J. L., Towns S. G., and Callaway C. B. Cosmo: A life-like animated pedagogical agent with deictic believability. In: Proc. of the IJCAI97 Workshop on Animated Interface Agents: Making them Intelligent, Nagoya, 1997
[2] Gratch J. Émile: Marshalling Passions in Training and Education. In: Proc. of Autonomous Agents 2000, 2000, 325–332.
[3] Ball G. and Breese J. Emotion and personality in a conversational agent. In [27], 189–219.
[4] de Carolis B., Pelachaud C., Poggi I., and Steedman M. APML, a Markup Language for Believable Behavior Generation. In [25], 65–85.
[5] Ortony A., Clore G. L., and Collins A. The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, MA, 1988
[6] Traum, D. R. and Rickel, J., Embodied agents for multi-party dialogue in immersive virtual worlds, In: Proc. of the First International Joint conference on Autonomous Agents and Multiagent systems, 2002, 766-773.
[7] Prendinger H., Saeyor S., and Ishizuka M. MPML and SCREAM: Scripting the Bodies and Minds of Life-Like Characters. In [26], 213–242.
[8] Lisetti C. L., Gmytrasiewicz, P., Can Rational Agent Afford To Be Affectless? A Formal Approach, Applied Artificial Intelligence, vol 16. 577-609, 2002
[9] El Jed, M., Pallamin, N., Dugdale, J., Pavard, B. Modelling character emotion in an interactive virtual environment. In: Proc. of AISB 2004 Symposium: Motion, Emotion and Cognition, 2004, Leeds, UK
[10] Wilson, I. The Artificial Emotion Engine, Driving Emotional Behavior, In: AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment, 2000
[11] Rousseau D., Hayes-Roth B. A Social-Psychological Model for Synthetic Actors, In: Proceedings of the 2nd International Conference on Autonomous Agents, 165-172, 1998
[12] Gebhard P., Kipp M., Klesen M., Rist T. Adding the Emotional Dimension to Scripting Character Dialogues In: Proc. of the 4th International Working Conference on Intelligent Virtual Agents, 2003, 48-56.
[13] Gebhard P., Klesen M., Rist T. Coloring Multi-Character Conversations through the Expression of Emotions. In: Proc. of the Tutorial and Research Workshop on Affective Dialogue Systems, 2004, 128-141.
[14] Davidson, R.J. On emotion, mood, and related affective constructs. In: P. Ekman & R.J. Davidson (Eds.) The Nature of Emotion: Fundamental Questions. New York: Oxford University Press. 1994, 51-55.
[15] Gebhard P. ALMA - A Layered Model of Affect. In: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, 2005, 29-36.
[16] Elliot C. Dissertation. The Affective Reasoner: A process model of emotions in a multi-agent system. Northwestern University, Evanston, IL, USA

[17] Gratch J., Marsella S. A Domain-independent Framework for Modeling Emotion. Journal of Cognitve Sytems Research, vol. 5, no. 4, 2004, 269-306.
[18] Ruttkay Z., Pelachaud C, From Brows To Trust, Evaluating Embodied Conversational Agents, Kluwer, Dordrecht, Boston, London, 2004
[19] Markus Löckelt. Action Planning for Virtual Human Performances. In: Proceedings of the International Conference on Virtual Storytelling 2005, Strasbourg, France, 2005
[20] Becker P. Structural and Relational Analyses of Emotion and Personality Traits. Zeitschrift für Differentielle und Diagnostische Psychologie, vol. 22, no. 3, 2001, 155-172.
[21] Morris W. N. Mood: The frame of mind. New York: Springer-Verlag, 1989
[22] McCrae R.R. and John O.P. An introduction to the five-factor model and its implications. Journal of Personality, vol. 60, 1992, 171–215.
[23] Mehrabian A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology, vol. 14, 1996, 261-292.
[24] Mehrabian A. Analysis of the Big-five Personality Factors in Terms of the PAD Temperament Model. Australian Journal of Psychology, vol. 48, no. 2, 1996, 86-92.
[25] Westermann, R. (1985). Empirical tests of scale type for individual ratings. Applied Psychological Measurement, 9, 265-274.
[26] Prendinger H. and Ishizuka, M. Life-Like Characters – Tools, Affective Functions, and Applications, Springer, 2004
[27] Cassell J., Sullivan J., Prevost S., and Churchill E. Embodied Conversational Agents. The MIT Press, Cambridge, Massachusetts, 2000