

# Coloring Multi-Character Conversations through the Expression of Emotions

Patrick Gebhard Martin Klesen Thomas Rist

DFKI GmbH  
Stuhlsatzenhausweg 3  
66123 Saarbrücken Germany  
{gebhard, klesen, rist}@dfki.de

**Abstract.** In this paper we describe how to generate affective dialogs for multiple virtual characters based on a combination of both automatically generated and pre-scripted scenes. This is done by using the same technique for emotion elicitation and computation that takes either input from the human author in the form of appraisal and dialog act tags or from a dialog planner in the form inferred emotion eliciting conditions. In either case, the system computes the resulting emotions and their intensities. Emotions are used to inform the selection of pre-scripted scenes and dialog strategies, and their surface realization. The approach has been integrated in two fully operable systems, the CrossTalk II installation and the NECA eShowroom.

## 1. Introduction

During the last years our group has explored the simulation of conversations among animated agents as a new style to present information to a user. A shift from settings with single presentation agents towards the use of presentation teams bears a number of advantages: They enrich the repertoire of modalities to convey information and they can serve as a rhetorical device to reinforce beliefs. In a first system, the so-called Inhabited Market Place [1], we focused on presenting information about a product, e.g., a certain car, in the form of a simulated dialog between a salesperson ( $S$ ) and one or more customers ( $C_1, C_2, C_3, \dots$ ). In essence, this system generated query-answer dialogs with queries referring either directly to product attributes or to value dimensions that are associated with the type of product under discussion. In the car domain the set of value dimensions includes prestige, safety, sportiness, environmental friendliness, etc. Typically, the role of virtual customers is to pose questions, whereas the virtual salesperson provides answers on the basis of available product information, e.g., product features stored in a database. In addition to asking questions or answering them, the virtual characters could also comment on a preceding turn of another character. Fig. 1 shows an excerpt of a generated car-sales dialog.

Regarding dialog generation the system has been implemented in two different variants. One variant deploys a single dialog planner that, somewhat similar to a playwright of a classical theatre performance, determines the course of the whole

conversation, i.e. all contributions of the involved characters together with the temporal order of who is going to speak next. In contrast to this, a second version of the system adopts the multi-agent paradigm and equips each of the involved characters with its own dialog planner [2]. While the multi-agent version is somewhat more challenging regarding turn-taking assignment and maintaining coherency of the conversation, it is more flexible as to the number of conversational partners and seems more suitable if the user herself takes part in the conversation as well. In either case, the generated dialogs are mostly task-oriented in the sense that the characters only talk about the presence or absence of certain product features, and the impact of features on associated value dimensions, in order to fulfill the underlying product information task.

S: *What can I do for you?*  
C<sub>1</sub>: *We are interested in that car.*  
C<sub>2</sub>: *How fast can it drive?*  
S: *It drives up to 225 km/h.*  
C<sub>2</sub>: *That's fast.*  
C<sub>3</sub>: *This is bad for the environment!*  
S: *Bad for the environment? This car is made of recyclable materials. Besides, it has a catalytic converter.*  
C<sub>1</sub>: *How much does it cost?*

Fig. 1: Portion of a simulated car-sales conversation

In order to make the simulated conversations more interesting for human observers, we wanted to insert sub-dialogs which would be perhaps off-topic but unexpected and funny. Unfortunately, the generation of such intermezzi is difficult to automate especially when compared to those that have been manually scripted by creative content authors. Therefore, we decided to develop a platform that enables to coherently interweave manual scripted sub-dialogs with dialogs that are automatically generated. This approach has been successfully applied for the realization of CrossTalk, an interactive installation designed for public spaces, such as an exhibition, or a trade fair [3,4].

The CrossTalk installation provides visitors with a spatially extended interaction experience by offering two virtual spaces on separate screens, one displaying Cyberella, the installation's hostess, the other displaying Tina and Ritchie, two virtual actors "hired" to perform car sales dialogs to inform visitors about a certain car. In addition, Tina and Ritchie can converse with Cyberella, thus creating the illusion that the agents have cross-screen conversations. This can be considered a playful illustration of the "computers as theatre" paradigm as introduced by [5]. Moreover, the installation relies on what we call a meta-theater metaphor. Quite similar to professional actors, characters in CrossTalk are not always on duty. Rather, they can step out of their roles, and amuse the user with unexpected intermezzi and rehearsal periods. For instance, when no visitor is present the installation is in OFF mode. To keep the actors alive for the purpose of attracting new passers-by, we emulate small-talk conversations among stand staff members. We rely on a repertoire of pre-scripted small-talk scenes from which the system would choose in OFF mode. For demonstrating CrossTalk at the CeBIT 2002 fair a large corpus of pre-scripted scenes (more than 220 for English and German each) has been assembled by one of our colleagues with experience in theater acting and directing. Some episodes cover themes related to every-day belongings, such as what to do in the evening, or to wonder about career prospects. Other scenes refer to the world of the theater or movies.

Fig. 2 provides an impression of a pre-scripted scene. In this case, the literal informational content is not really of interest to a third-party observer while the reflection of affective state and interpersonal relationships among the characters may be. In fact, informal feedback collected from visitors who interacted with the installation and watched performances of the characters revealed that almost all of them found the scripted episodes more amusing and interesting to watch than the automatically generated car-sales performances.

Cy.: <i>What are you gonna do after this?</i>
Tina: <i>Got a job at Walmart.</i>
Cy.: <i>Online sales?</i>
Tina: <i>Something like that.</i>
Ritchie: <i>You're kidding. You're gonna be a sales puppet?</i>

Fig. 2: Pre-scripted scene

This observation motivates the aim to color simulated dialogs through the expression of emotions in order to create a lively performance. To this end, we first need to increase the cognitive model of our characters by allowing them to have emotions. Secondly, we need to simulate changes in emotions as consequences of executed dialog acts. Thirdly, we need to refine dialog generation so that emotions are taken into account and exploited in the generation process.

## 2. Related Research

Virtual characters and especially embodied conversational characters are now widely used in various application areas including virtual training environments [6], portable personal guides [7], interactive fiction [8] and storytelling systems [9], as well as e-commerce applications [10], and in interfaces of consumer electronics [11].

Many conversational characters have been developed for applications assuming that the character engages in a face-to-face conversation with the user. In order to increase the believability of the virtual conversation partner, researchers have begun to address the modeling and emulation of human-like qualities such as personality and affective behavior. Examples of emotional characters include COSMO [12], Émile [13], Peedy [14], and the Greta agent [15]. In these systems emotion modeling has been inspired by the so-called *OCC* model developed by Ortony, Clore, and Collins [16], although the approaches differ in the granularity of modeling, the mathematical machinery for computing emotions, and in the way of how the model has been implemented on a technical level.

Systems with multiple characters have also been proposed by others. Earlier approaches, such as Gilbert & George [17], Mr. Bengo [18], and [19], however, they do not explicitly model affective states in their characters. In contrast to these approaches, 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 [20] address dialog management comprising human-character and character-character dialogs in immersive virtual environments. Prendinger et. al. [21] 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 the characters. The simulation of group dynamic

phenomena in multi-character scenarios has been addressed by [22] and [23]. Both approaches are based on socio-psychological theories, Guye-Vuillieme et. al. aim at the generation of believable displays of non-verbal interactions while Rist and Schmitt simulate attitude changes of virtual characters in multi-party negotiation dialogs. Finally, the work by [24] is of high relevance for our work since it addresses the variation in linguistic style with regard to some socially determined variables. However, their generation approach does not distinguish between emotion categories but represents a character’s affective state just by a single abstract parameter called emotional disposition.

### 3. Outline of Approach

As stated in the introduction, our aim is to improve the quality of simulated conversations among virtual characters by modeling their affective states in order to exploit such states as an additional resource for the generation of dialog contributions, the articulation of verbal utterances, and the generation of non-verbal expressions.

#### 3.1 Modeling emotions and personality traits

To give our characters the ability to have emotions, we follow the lines of the *OCC* model [16]. OCC is 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 us, actions of those we consider responsible for such actions, and objects/persons. The OCC theory defines 22 emotion types. For the purpose of the current paper, we concentrate on those listed in Table 1.

Table 1: Examples of OCC-Emotions

Group	Description	Emotion Type and Name
Well-being	Appraisal of a situation as an event.	Joy: an event is desirable for self. Distress: an event is undesirable for self.
Prospect-based	Appraisal of a situation as a prospective event.	Hope: a prospective event is desirable. Fear: a prospective event is undesirable.
Attribution	Appraisal of a situation as an accountable action of some agent.	Pride: approving of one’s own action. Admiration: approving of another’s action. Shame: disapproving of one’s own action. Reproach: disapproving of another’s action.
Attraction	Appraisal of a situation as containing an attractive or unattractive object.	Liking: finding an object appealing. Disliking: finding an object unappealing.

Emotion intensities and their decay differs across personalities. Similar to other approaches, we rely on so-called personality dimensions (or traits) to describe the personality of our characters. In particular, we consider the personality traits proposed in the Five Factor Model of personality [25]: openness, conscientiousness, neuroticism, extraversion, and agreeableness. We represent a character’s conformance with a certain trait as an adjustments on a scale between  $-1.0$  (not all conformant) and  $+1.0$  (fully conformant). This way, the personality profile of a character is given by the set of adjustments for all traits. Depending on the degree to which a character conforms with a certain trait both intensity of emotional reactions and decay of emotion intensities are affected. For instance, we model our characters to feel a *joy* emotion more intensely the more extravert and agreeable they are. We express dependencies of this kind by biasing the emotions with a baseline intensity according

to the settings of personality traits, e.g. an extravert character's baseline intensity for joy is 0.15, whereas an introvert character's baseline intensity for joy would be 0.0. We use the openness and conscientiousness traits to influence the emotions intensity in the attribution and attraction group. For instance, a conscientious character appraises the result of an irresponsible behavior as more blameworthy than a less conscientious character.

Fig. 3 shows a graphical interface that can be used to specify a character's personality traits (down right), to choose among several emotion decay functions (top right), and an emotion monitor that traces intensity development and decay of emotions during a simulated conversation (left). A check mark on the right hand (see hope) signals that this emotion currently impacts the character's non-verbal and verbal behavior. Furthermore the elicitors of an emotion are displayed in an hover frame. For testing/redesigning purposes this interface allows to manipulate a character's emotions at run time (second slider to the right of each emotion).

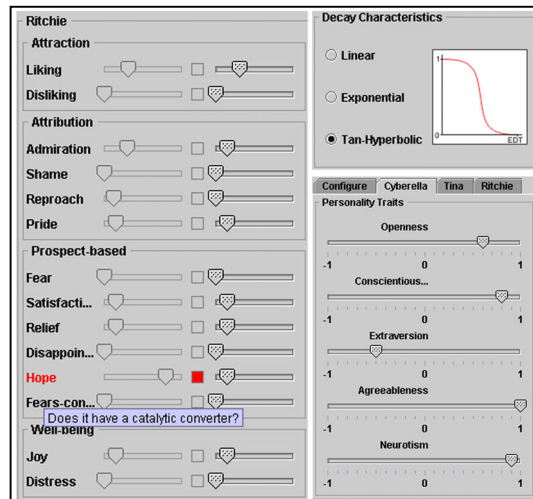


Fig. 3: Emotion monitor and configuration panels.

For testing/redesigning purposes this interface allows to manipulate a character's emotions at run time (second slider to the right of each emotion).

### 3.2 Emotion elicitation and change of affective states

A central aspect of the OCC model is that it associates each emotion type with a set of emotion eliciting conditions (EECs). We represent EECs as a set of variables that characterize emotion-inducing situations. We currently use the following set of variables: *Desirability* of events, *Praiseworthiness* of actions, *Appealingness* of objects, *Liking* reflecting how attracted to another person one is, *Likelihood* reflecting the degree of belief that an anticipated event will occur, and *Realization* reflecting the degree of belief that an anticipated event has occurred. EEC variables take on values from the range [-1.0, 1.0]. For example, in the case of *Desirability* the value -1.0 means *very undesirable* (induces *distress*) whereas 1.0 stands for *very desirable* (induces *joy*). The signs of the values determine which emotion types are elicited (e.g. *distress* or *joy*) and the absolute values determine their intensities. The outcome of the subjective appraisal is a list of emotions:  $(e_1, \dots, e_n)$ .

Applying OCC to compute emotions in dialogs among virtual characters starts by identifying what kind of events, actions and objects can occur and how to map them to EECs. Our characters live in a virtual (2D or 3D) environment and have limited perceptual capabilities. In the CrossTalk installation, they are only able to recognize a limited number of events, such as user feedback via touch screen, simulated system failures, or the arrival and departure of visitors detected by a web cam. Since we focus on group conversations, we can circumvent this limitation. Our characters learn about

events, actions, and objects by listening to the dialog contributions of the other conversational partners. We hereby assume that the internalization of events, actions or objects is obtained either through a direct perceptual process or results from interpreting dialog acts. In either case, we associate events, actions or objects with EECs. Since we deal with multi-character scenarios, we need to model that different characters may perceive the same event, action or object differently, and in case of overhearing dialog contributions (referring to events, actions or objects) they may develop different emotional responses. Therefore, events, actions or objects must be mapped to ECCs separately for each character. In sum, our model comprises several mapping steps as illustrated in Fig. 4.

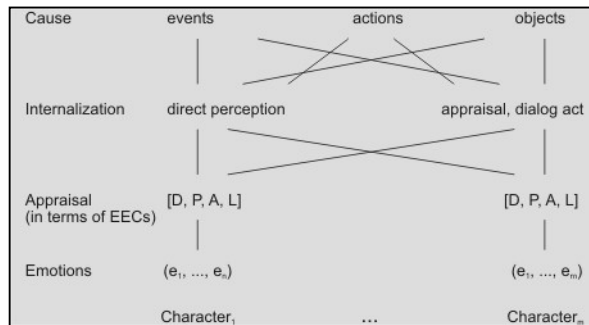


Fig. 4: Steps in determining emotions

#### 4. Generating Affective Car-Sales Dialogs – the NECA eShowroom

The NECA eShowroom ([www.eshowroom.org](http://www.eshowroom.org)) is a more elaborated version of the Inhabited Market Place [1] introduced in Section 1. Based on user input the system generates an animated car sales dialog between virtual characters [26]. The eShowroom consists of the following main components: a dialog planner, a multimodal natural language generator, a text/concept-to-speech synthesis, and a gesture assignment module.

The dialog planner takes the role of a playwright that generates a script (a sequence of dialog acts) for the virtual characters that become the actors in a simulated car sales dialog. A dialog act represents a communicative function, such as requesting some information (request), answering a question (confirm, inform), or giving feedback (agree). Dialog strategies represent a certain sequence of dialog acts (or moves) of the participating dialog partners as they can be typically observed in the genre at hand. A sales dialog typically starts with a greeting phase, followed by the customer's request for information about a specific product. Subsequently, a question-answer game between the customer and the salesperson develops where various aspects of the product are discussed. Finally, the customer communicates a purchase decision and, in a closing phase, the dialog ends.

The generated script that encompasses all dialog moves is then handed over to the multimodal natural language generator, which transforms dialog act representations into text, annotated with syntactic, semantic, and pragmatic features. The component is also responsible for nonverbal behavior, such as selecting iconic or emblematic gestures. The task of the text/concept-to-speech synthesis is then to convey, through adequate voice quality and prosody, the intended meaning of the text as well as the emotion with which it is uttered. It also provides information on the exact timing of

utterances, syllables and phonemes, which is indispensable for the gesture assignment module. This module is responsible for fine-tuning the synchronization of speech and gesture, including proper alignment of verbal and nonverbal output.

For generating affective dialogs among virtual characters we make two extensions to our dialog planner. Firstly, we augment the dialog act representation so that it comprises an emotion slot. It will be instantiated by the most dominant emotion of the act performer (i.e., the speaker).

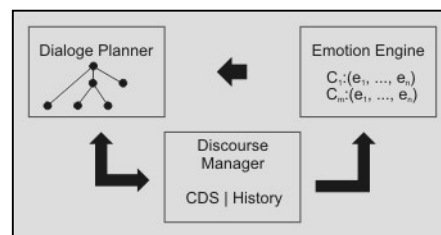
Secondly, we need to interleave dialog planning with our mechanism for emotion elicitation and computation. In our approach this coupling is done via the concept of a current discourse state (*CDS*) as illustrated in Fig. 5. The current discourse state depends on context information, such as knowledge concerning the dialog partners and their actions, as well as events and objects in the world. When a dialog act is performed, the *CDS* is updated. Since the dialog act may have an impact on the characters' emotions, a so-called Emotion Engine performs an update of the characters' affective states. In other words, we compute how a new situation is

appraised by the dialog partners. When the dialog planner determines the next dialog move, i.e., speaker and dialog act, it will take into account the updated affective states. In particular, the speaker's most dominant emotion will be used as an additional parameter for text generation, gesture alignment, and speech synthesis. The performance of the dialog act will cause a next update of the *CDS* and in turn change the affective states of the characters.

For the purpose of illustration let us assume that a virtual salesperson has just introduced a new car to the customer and said: "This is a very sporty car, it has 240 horse powers". This new information about the existence of an object with this feature will be appraised by the customer, either positively or negatively depending on the customer's attitudes towards sporty cars. If positively appraised, a liking emotion will be elicited and depending on its intensity, it may be reflected in the customer's next dialog move, e.g. "Wow, 240 horse powers, that's wonderful!"

Somewhat more tricky are question-answering games. Suppose the customer would like to continue the dialog by posing a question about another car attribute. To this end, the repertoire of dialog strategies comprises strategies of the type "QuestionAnswer:Boolean" in which the customer requests information about a boolean attribute, e.g. by asking whether the car has a catalytic converter. Depending on the attribute value, the salesperson will confirm or disconfirm this.

In order to be able to reflect emotion in the query, e.g. through a modification of the voice quality, we need to anticipate the appraisal of the situation in which the car actually has the requested feature. Suppose the customer pays attention to consumption and environmental friendliness. The presence of a catalytic converter will then be appraised as highly desirable ( $D = 0.8$ ). However, since at the current state in the dialog this is not yet confirmed the odds may be fifty-fifty so that the EEC variable Likelihood takes on the value 0.5 ( $L := 0.5$ ). Based on these EECs a hope emotion will be elicited with a relatively high intensity (hope := 0.82). Assuming that



**Fig. 5:** Coupling a Dialog Planner with an Emotion Engine via a Discourse Manager

this is the most dominant emotion in the current situation, the dialog planner will assign the hope emotion to the request act. At the surface level, it may then be reflected in the formulation of the query, e.g., by using the wording:

*“It certainly does have a catalytic converter, doesn’t it?”*

accompanied by a hopeful facial expression.

## **5. Interweaving Emotional Scripting and Dialog Generation – CrossTalk II**

As mentioned in Section 1, a peculiarity of the CrossTalk installation is that dialog simulation relies partly on a plan-based generation approach and partly on a repertoire of human authored scenes. To design an new version of CrossTalk with affective characters (CrossTalk II), we use the plan-based approach for affective dialog planning as sketched in Section 4. As in the original version of CrossTalk, a simple authoring syntax is meant to appeal to non-programmers, including professional playwrights, and scripts can be written with any text processing software. To be able to interweave authored and automatically generated scenes at runtime, a script compiler transforms authored scenes into the same internal format that is used by the dialog planner [3,4].

Since we now model emotions in our characters, we need to make sure that authored dialog contributions have an impact on the emotions, too. In principle, one could try to automatically extract emotion eliciting conditions for the characters by analyzing and interpreting the authored scene scripts. Another approach is to provide the authors with a convenient method to add meta-level information relevant to the eliciting of emotions. In CrossTalk II we follow this approach and provide scene authors with two kinds of tag sets: *appraisal tags* and *dialog act tags*, see [27].

### **5.1 Appraisal Tags**

*Appraisal Tags (ATs)* express how a speaking character appraises the event, action or object about which it is talking. ATs serve as symbolic abbreviations for EECs to release scene authors from the burden of specifying EEC variables. Using brackets to indicate tags in a scene script, appraisal tags are inserted directly behind the transcriptions of dialog contributions as shown in the example below:

R: *The weather’s getting better.* [=good\_likely\_future\_event]

When processing a script, CrossTalk’s scene compiler maps ATs onto EECs of the speaking character by using default settings for the EEC variables. For instance, the appraisal tag in the example above is mapped to the EEC variables:

[D:= 0.5 (moderately desirable), P:= nil, A:= nil, L:= 0.5 (moderately likely)]

### **5.2 Dialog Act Tags**

*Dialog Act Tags (DATs)* are somewhat more specific annotations than ATs. Their function is to enable human scene authors with an intuitive means to indicate the underlying communicative intent of an utterance. The tags are inserted behind the



utterance they refer to. The dialog act's addressee(s) must be specified. For example, Figure 6 shows the use of a dialog act tag in an authored scene. The utterance of the character Ritchie is a verbal attack on the character Tina. For CrossTalk II we have defined currently 23 DATs and an addressee can be one of the characters Tina (T), Ritchie (R), Cyberella (C), the user (U) or all conversation partners (ALL).

```
Tina: I didn't get the job for the MTV
webpage. It went to some kid that
looked like Britney Spears.
Ritchie: Well, can you sing? [=attack T]
```

Fig. 6: Authored scene with an dialog act tag

Similar as in the case of appraisal tags, the Scene Compiler maps DATs onto EECs. However, in contrast to appraisal tags which concern only the EECs of speakers, DATs concern speaker, addressee, and in some special cases even further characters not explicitly specified as addressees in the DAT. For instance, in CrossTalk the character Cyberella has the role of a moderator. Therefore, if the two actors Tina and Ritchie attack each others in front of the audience, she may consider this as an undesirable event, too. To release the script author from the burden of tagging all dialog moves in full detail, however, the author can define specific mapping rules for each DAT. Regarding the example in Figure 6, the corresponding mapping of the dialog act tag *attack* is shown in Figure 7.

```
$Speaker $Text [=attack $Addressee] ::=
THEN ADD $Addressee-EEC:
D:= -1.0 (very undesirable)
P:= -1.0 (very blameworthy)
THEN ADD $Speaker-EEC:
D:= +0.5 (moderately desirable)
AND IF $Speaker NOT Cyberella
THEN ADD Cyberella-EEC:
D:= -0.5 (moderately undesirable)
P:= -0.5 (moderately blameworthy)
```

Fig. 7: Example of mapping DAT to EECs

### 5.3 Tagging user feedback

In the NECA-eShowroom as well as in CrossTalk the user is primarily in the role of spectator who cannot participate in the conversations of the virtual characters. CrossTalk, however, solicits user feedback on its performances. When the characters Tina and Ritchie act out a car-sales performance, the user's touch screen displays an *applause* and *boo* button that can be pressed anytime in order to indicate appreciation or dislike. We treat such indications similar to dialog acts that are directed to the characters and that have a certain impact on their emotions. More technically speaking, each button is associated with a dialog act tag which is in turn mapped onto EECs and eventually cause an update of the characters' emotions. This way CrossTalk characters react emotionally on user input. Thus, by giving frequent feedback the visitor can influence the emotions of the characters and elicit emotional responses. For instance, in case a visitor keeps on pressing the boo button (negative feedback) for a certain period of time, Cyberella becomes more and more *distressed*. Her emotional disposition may eventually change her dialog behavior. Depending on the current state of a car-sales performance she may interrupt and suggest to start again after the user has provided new parameter settings.

## 6. Exploiting Affect in Dialog Generation

There are many decision points in our applications at which knowledge about the emotions of speakers and addressees can be taken into account. Focusing on multi-party conversations (rather than performing physical actions), emotions can be used to inform the selection of dialog strategies and linguistic style strategies as proposed by [24]. 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 by selecting the dialog act type and the propositional content. For each dialog act the system needs then to make a linguistic choice regarding wording and surface forms. In the speech production system the emotions should be reflected in the intonation and voice quality. In addition, the non-verbal behavior of both speaker and listeners needs to be considered. This includes gaze behavior, facial display of emotions, postures, gestures, and proxemics, i.e. maintaining interpersonal distances and interaction angles [28].

In the NECA eShowroom (Section 4) knowledge about the emotions of the dialog participants is used by the dialog planner, the multimodal natural language generator (M-NLG), and the speech synthesis module. The dialog planner evaluates the preconditions of the dialog strategies and selects the one that best matches the affective states of the characters. For instance, if the salesperson repeatedly says “I don’t know.” when being asked by the customer about certain features, the customer will get more and more frustrated. If the anger increases a certain threshold the dialog planner interrupts the ongoing question-answer game and selects a new dialog strategy that leads over to the closing phase. The M-NLG takes the emotion into account when creating the surface realization of a dialog act. Depending on the speaker’s emotion, the dialog act “inform about interior” could be realized as “It has a cramped interior.” (disliking) versus “The interior is more spacious than you would expect.” (liking). The selection of communicative gestures and facial expressions is also informed by the specified emotion. For this purpose we map emotions to the available facial animations (morph targets) and use the emotion intensity to determine the intensity of the facial expression. Our speech synthesis module works with so-called *emotion dimensions*, a simplified representation of the essential properties of emotions [29]. In NECA we have defined a mapping from OCC emotion types and intensities to the three emotion dimensions *evaluation* (positive vs. negative), *activation* (active vs. passive), and *power* (dominant vs. submissive). The dimensional approach to emotional speech synthesis uses rules to map any point in this three-dimensional emotion space onto its acoustic correlates. The resulting synthesis system allows the gradual build-up of emotions, the synthesis of non-extreme emotions, and the change of emotional tone over time [30].

In the CrossTalk II system (Section 5) the emotions of the virtual actors are used by the presentation manager, the output renderer, and the speech synthesis module. CrossTalk’s central module, the *presentation manager*, is responsible for selecting and executing scenes. Which scene is played next is defined in the scene flow. Emotions are used at runtime when making transitions (conditional branching) and when selecting between alternative scenes. For instance, when the user gives feedback during a performance, a feedback scene is played which depends on the type of feedback (applause or boo) and on the affective state of the two virtual actors. The

presentation manager also contains a dialog planner, which generates scenes at runtime. It uses affective states in a same way as they as in the NECA eShowroom.

The presentation manager forwards directions for virtual characters contained in scenes to the *output renderer* module. These directions are mapped to character actions and input for the speech synthesis. At this level, we use emotions to enrich the dialog contributions provided by the human author. We constantly change the characters' facial expressions so that they reflect their emotions. This is done in the same way as in the eShowroom, i.e. by mapping emotions and their intensities on the available facial animations. In some cases gestures are triggered to reflect a characters emotions, e.g. if Ritchie is angry with Tina, he will automatically use a pointing gesture at Tina, when he makes a verbal attack.

In CrossTalk we use commercial text-to-speech system for generating speech output. It provides an interface for changing speech parameters like baseline pitch, speed, and volume. Using rules of thumb and intuition we have established relationships between emotions and speech parameters, for example, to express joy, we raise the baseline pitch and the volume according to the emotion intensity [27].

## 7. Discussion and Future Work

In this paper we have reported on our approach to add affect to simulated conversations among virtual characters. We started off from two concrete systems that generate dialogs but none of which did explicitly model emotions in the characters. We first extended our characters towards affective virtual personalities along the lines of the OCC model. By means of the so-called NECA eShowroom system we showed how a plan-based approach to dialog generation can be coupled with the computation of emotions. In our second system, CrossTalk II, we further showed how to interweave plan-based generation with human authoring of individual scenes to produce coherent affective dialogs. To this end we have introduced special tag sets that allow human scene authors to make annotations that impact the emotions of scripted characters. Abstracting from details on the syntactical level, this approach shares similarities with proposals for character scripting languages comprising emotion tags, such as AML, APML, MPML, or VHML [31]. However, a peculiarity of our approach is that we use the same machinery for emotion elicitation and computation for both automated dialog generation and the execution of human authored scenes. This allows us to smoothly interweave both approaches at the emotional level. The outcome of our work comprises two fully operable systems, the CrossTalk II installation, and the NECA eShowroom. Still, there is much room for further improvements.

We have made quite a number of simplifications regarding the implementation of OCC. We are aware of the fact that more elaborated OCC implementations for virtual characters exist that we could have adopted in principle. We preferred to start with a more simplistic approach since in both of our applications we deal with multiple characters, and in the case of CrossTalk II we had to address the additional integration of plan-based generation with human authoring. Having two running systems at hand

will allow us to rapidly implement refinements of the emotion models and test their impact on the quality of the emerging conversations.

Another yet not sufficiently addressed aspect concerns the exploitation of emotions in the generation of multi-modal dialog behaviors in a multi-party setting. In the context of the VirtualHuman project we will use knowledge about emotions more rigorously to refine decision making in our characters regarding action selection, dialog planning, and multimodal articulation.

Finally, on a longer term perspective it would be desirable to have users fully participate in affective conversations. However, this requires much more research on reliable analysis and interpretation methods for processing spoken input and inferring the user's affective state.

## Acknowledgements

This work has been built upon contributions from the VirtualHuman project ([www.virtual-human.org](http://www.virtual-human.org)) funded by the German Ministry for Education and Research and from the EU-funded IST NECA project.

## References

- [1] André E., Rist T., van Mulken S., Klesen M., and Baldes S.. The automated design of believable dialogues for animated presentation teams. In [32], pages 220–255.
- [2] Rist T, André E, Baldes S In: Proc of IUT'03, A Flexible Platform for Building Applications with Life-Like Characters ACM Press, 2003, pp 158-165.
- [3] Klesen M., Kipp M., Gebhard P., Rist T. Staging Exhibitions: Methods and tools for modelling narrative structure to produce interactive performances with virtual actors, *Journal of Virtual Reality – Special issue on Storytelling in Virtual Environments*, 2003.
- [4] Gebhard P., Kipp M., Klesen M., Rist T. Authoring Scenes for Adaptive, Interactive Performances, In: *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-03)*, 2003.
- [5] Laurel, B. *Computers as theatre*. Addison-Wesley, Reading MA, 1993
- [6] Swartout, W., Hill, R., Gratch, J., Johnson, W.L., Kyriakakis, C., LaBore, C., Lindheim, R., Marsella, S., Miraglia, D., Moore, B., Morie J., Rickel, J., Thiébaux, M., Tuch, L., Whitney, R., and Douglas, J. Towards the Holodeck: Integrating Graphics, Sound, Character and Story. *Proc. of Autonomous Agents'01*, 2001, 409-416
- [7] Rocchi C., Stock O., Zancanaro M., Kruppa M., and Krüger A. The Museum Visit: Generating Seamless Personalized Presentations on Multiple Devices In: *Proc. of Intelligent User Interfaces 2004*.
- [8] Murray, J. H. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. The MIT Press, 2000.
- [9] Ryokai, K., Vaucelle, C., Cassell, J. Virtual peers as partners in storytelling and literacy learning. In: *Journal of Computer Assisted Learning*, 19(2), 2003, pp. 195-208
- [10] McBreen H, Jack M Evaluating Humanoid Synthetic Agents in E-Retail Applications *IEEE Trans. On Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol. 31, N0. 5, 2001, pp. 394-405
- [11] Diederiks E. Buddies in a Box Animated Characters in Consumer Electronics. In *Proc. of IUT'03*, 2003, pp. 34-38

- [12] 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
- [13] Gratch J. Émile: Marshalling Passions in Training and Education. In Proc. of Autonomous Agents 2000, pp 325–332
- [14] Gene Ball and Jack Breese. Emotion and personality in a conversational agent. In [32], pages 189–219.
- [15] de Carolis B., Pelachaud C., Poggi I., and Steedman M. APML, a Markup Language for Believable Behavior Generation. In [31], pages 65–85.
- [16] Ortony A., Clore G. L., and Collins A. The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, MA, 1988.
- [17] Cassell J., Pelachaud C., Badler N., Steedman M., Achorn B., Becket T., Douville B., Prevost S., Stone M., Animated Conversation: Rule-Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. In: Proc. of SIGGRAPH '94, 1994, pp. 413-420.
- [18] Nitta K., Hasegawa O., Akiba T., Kamishima T., Kurita T., Hayamizu S., Itoh K., Ishizuka M., Dohi H., and Okamura M. An experimental multimodal disputation system. In: Proc. of the IJCAI '97 Workshop on Intelligent Multimodal Systems, 1997.
- [19] Rickel J, Johnson W. L. Virtual Humans for Team Training in Virtual Reality. In: Proc of the Ninth International Conference on Artificial Intelligence in Education,. Amsterdam: IOS Press., 1999, pp 578-585
- [20] 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.
- [21] Prendinger H., Saeyor S., and Ishizuka M. MPML and SCREAM: Scripting the Bodies and Minds of Life-Like Characters. In [31], pages 213–242.
- [22] Guye-Vuillieme A., Thalmann D. A High Level Architecture For Believable Social Agents, VR Journal, Springer, Vol.5, 2001, pp.95–106.
- [23] Rist T., Schmitt M. Applying Socio-Psychological Concepts of Cognitive Consistency to Negotiation Dialog Scenarios with Embodied Conversational Characters. In: Proc. of AISB'02 Symposium on Animated Expressive Characters for Social Interactions, 2002, pp79–84.
- [24] Walker M., Cahn J., and Whittaker S. J. Improving linguistic style: Social and affective bases for agent personality. In: Proc. of the First International Conference on Autonomous Agents, 1997, pp. 96–105.
- [25] McCrae R.R. and John O.P.. An introduction to the five-factor model and its implications. *Journal of Personality*, 60:171–215, 1992.
- [26] Krenn B., Pirker H., Grice M., Piwek P., van Deemter K., Schröder M., Klesen M. Generation of multimodal dialogue for net environments. In: Proc. der 6. Konferenz zur Verarbeitung natürlicher Sprache, 2002.
- [27] 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 (IVA'03), Kloster Irsee, 2003.
- [28] Kendon A. Nonverbal Communication, Interaction, and Gesture. Mouton, The Hague., 1980
- [29] Russel, J. A. How shall an emotion be called. In: Plutchik, R. and Conte H. R. (eds): Circumplex models of personality and emotions, American Psychological Association, 1997
- [30] Schröder M. Speech and Emotion Research: An overview of research frameworks and a dimensional approach to emotional speech synthesis. PhD thesis, Institute of Phonetics, Saarland University, 2003
- [31] Prendinger H., Ishizuka, M. Life-Like Characters – Tools, Affective Functions, and Applications, Springer, 2004
- [32] Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill, editors. Embodied Conversational Agents. The MIT Press, Cambridge, Massachusetts, 2000.