Adding the Emotional Dimension to Scripting Character Dialogues

Patrick Gebhard Michael Kipp Martin Klesen Thomas Rist

DFKI GmbH
Stuhlsatzenhausweg 3
66123 Saarbrücken Germany
+49 681 302 5152
{gebhard, kipp, klesen, rist}@dfki.de

Abstract. We present an extension of the CrossTalk system that allows to model emotional behaviour on three levels: scripting, processing and expression. CrossTalk is a self-explaining virtual character exhibition for public spaces. Its SceneMaker authoring suite provides authors with a screenplay-like language for scripting character and user interactions. This article presents an extension to the original CrossTalk scripting language by providing a set of appraisal and dialogue act tags, making emotional behaviour generation possible. These extensions rely on CrossTalk's new EmotionEngine which processes emotions by computing and maintaining emotional states for each character. In combination with the ContextMemory module it enables the characters to adapt to user feedback and to react to previous encounters with users in an emotional way. We describe the use of the appraisal and dialogue act tags, their processing in the EmotionEngine and their impact on the characters' verbal and non-verbal expressive behaviour.

1. Introduction

Animated conversational characters are widely used in various application areas. This includes virtual training environments [1, 2], interactive fiction [3, 4] and storytelling systems [5], as well as e-commerce applications.
The use of multiple characters in the various applications allow to convey social aspects such as interpersonal relationships between emotional characters [6, 7]. Designing believable behaviour can become a complex task, especially if multiple characters are involved and user interaction has to be taken into account. Characters must respond in a way both appropriate and non-repetitive because otherwise their believability is undermined. The major question is: where does the script come from which defines the verbal and nonverbal behaviour? There are basically two approaches: The system can play the role of a playwright that automatically generates scenes at runtime [8] or the system uses pre-scripted scenes authored by a human writer. In [9] we have described an approach that combines human scripting with automated script generation. Its central paradigm is the separation of narrative

structure, the scene flow, from story content which, from an author's viewpoint, facilitates controlling story structure and consistency.
For creating believable behaviour there is now a common agreement that emotions and personality are key ingredients [10, 11, 12, 13]. Emotions seem to be a good mediator between actions, events and objects on one hand and behaviour on the other hand. We introduce dialogue act tags to represent actions in the script and appraisal tags to represent events. These tags together with context knowledge, enable CrossTalk's EmotionEngine to generate and maintain emotions for each character. These emotions trigger appropriate verbal and non-verbal behaviour and let the characters react in user interaction in an emotional way.
The presented approach of using scripting tags for computing emotions is related to the Multimodal Presentation Markup Language (MPML) tags used for the SCREAM system [14]. SCREAM is comparable to CrossTalk's EmotionEngine, but differs in emotion processing and maintaining. MPML was created for scripting presentations with multiple characters in a web-based environment. Although it does not separate content and scene flow like CrossTalk, it provides tags from which character emotions are computed. While it is possible to control the behaviour of multiple characters, it does not generate emotional expression in speech, a point that will be addressed in this paper. Other related scripting approaches are the Virtual Human Markup Language (VHML) [15] or the Affective Presentation Markup Language (APML) [16]. Most of these languages concentrate on scripting locomotive and expressive behaviour like facial expression, body animation, and other control tags. Due to the fact that they mostly rely on low-level control mechanisms for scripting character behaviour, they are difficult to use for authors without programming skills.
The article is organized as follows. Section 2 describes the CrossTalk installation and the underlying software framework. Sections 3-6 describe the use of emotional markups in CrossTalk's dialogue scripting language to compute emotions. Section 7 addresses the automatic behaviour generation based on emotion and context information. Finally, we give a conclusion and present some related future work.

2. CrossTalk

CrossTalk is a 24-hour interactive entertainment installation with virtual characters that are "alive" around the clock. It 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. Together with the user console up front it creates an interaction triangle. The installation needs no personnel – it is self-explaining and runs in an endless loop. The CrossTalk installation has two major operation modes: ON mode and OFF Mode. When no user is present the installation is in OFF mode. In this mode, passers-by are attracted by the ongoing "life" in the background screens, while the third front screen features a large, pulsating "start" button. In OFF mode, all three characters perform idle actions like breathing, looking around or shifting posture. They also chat with each other and even start rehearsing parts of the performance from time to time. As soon as a passer-by enters the installation by pressing the

"start" button, the application switches to ON mode. All characters interrupt their current activities and Cyberella welcomes the user and offers a demo performance of an automatically generated car sales dialogue. She guides the user through parameter selections where the user can choose roles, personalities and issues for the dialogue acted out by Ritchie and Tina. During the car sales dialogue, the user's screen shows three feedback buttons that read "applause", "boo" and "?". If the user pushes one of these, an adequate reaction is interweaved into the running demo. The underlying design principles and a full system description are explained in [17].
CrossTalk's three separated screens are driven by three computers that are integrated by a distributed software architecture. The central module, the PresentationManager, controls all virtual actors and handles user interaction. It selects and executes scenes. The ContextMemory module stores user and system actions and interactions (discourse history) and situational context (e.g., day time or location). The context memory regards the agents' contributions as well as the user feedback as utterances, storing speaker, addressee and content. The EmotionEngine module computes an emotional state for each character (see Section 6). This is primarily done by a context dependent appraisal of relevant events, actions and objects. The context information, like speaker, addressee, etc. is provided by the ContextMemory. At the end of the generation pipeline, the OutputRenderer module maps the internal character command representation on character player specific commands. This includes the mapping of emotion representations on related expressions, such like speech parameters and gesture commands. For the character rendering, we currently rely on the Microsoft Agent player technology [18].

3. The Emotional Dimension

The use of emotions and their expression in verbal and non-verbal behaviour requires a computational model of emotions. We use a cognitive model of emotions that maps external events, actions and the attitude towards objects/persons onto emotional states including a measure of intensity. The CrossTalk implementation of this model is the EmotionEngine module. It allows to reason about characters' emotional states based on *emotion eliciting conditions* (*EECs*) for twenty-four emotion types using the so-called *OCC* model developed by Ortony, Clore, and Collins [19].
Processing emotions starts by assessing which aspects of the world can trigger emotions. Three aspects can be distinguished: *events* of concern to us, *actions* of those we consider responsible for such actions, and *objects*. The cognitive appraisal of these aspects is based on a number of factors like the agent's goals, standards, attitudes and results in an *emotion eliciting condition* (EEC) that describes the desirability of the event, the praiseworthiness of an action and the appealingness of an object. Appraisal can be performed using a table which directly maps events, actions and object onto EECs, or doing complex reasoning using the agent's internal state and world knowledge. The OCC model uses EECs to compute emotions and their intensities.
In CrossTalk, in order to derive EECs, we introduce *dialogue act tags* and *appraisal tags* in the scripting language. An appraisal tag is an abbreviation for a specific EEC and can be directly processed by the EmotionEngine. Dialogue act tags implicitly

model a number of EECs and can easily be annotated by a human author. They are interpreted to contain aspects of events, actions and objects. The following sections describe the appraisal and dialogue act tags of the CrossTalk dialogue scripting language, how they are mapped onto EECs for the computation of emotions and finally, the expression of emotions.

4. Emotion Eliciting Conditions

Appraisal tags express how a character appraises an event, action or object. The tags are inserted behind the utterance:

R*: The weather's getting better*. [=good_likely_future_event]

Appraisal tags are an abbreviation for a set of EECs. Adding EEC descriptions into a dialogue script would interfere with the approach of keeping the annotation as lean as possible. Table 1 shows all available appraisal tags. To infer emotional states the EmotionEngine needs to know how to map these tags onto the EEC variables *desirability* (*D*), *praiseworthiness* (*P*), *appealingness* (*A*) and *likelihood* (*L*). This mapping is represented as a lookup table using default values for the degree of the EEC variables, e.g. the appraisal tag [=good_likely_future_event] in the example above is mapped to:

Table 1: Appraisal tags.

Class	Appraisal tag	
event	good_event	bad_event
	good_unlikely_future_event	bad_unlikely_future_event
	good_likely_future_event	bad_likely_future_event
action	good_act_self	bad_act_self
	good_act_other	bad_act_other
object	nice_thing	nasty_thing

D: +0.5 (moderately desirable) L: +0.5 (moderately likely)
A: nil (no impact) P: nil (no impact)

EEC variables have degrees ranging from –1.0 (e.g. very undesirable) to +1.0 (e.g. very desirable). The variables determine the computed emotion whereas the variables' degrees influence the emotional intensity.

5. Dialogue Acts as Events and Actions

Dialogue act tags define a rudimentary semantic for the utterance. It expresses the speaker's communicative intent. The tags are inserted behind the utterance they refer to. The dialogue act's addressee must be specified, too. For example:

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

The addressee can be the characters Tina (T), Ritchie (R), Cyberella (C), the user (U) or all (ALL). Currently, we have specified 23 dialogue acts (see Table 2).

For emotion computation, the dialogue act tags are mapped to EECs. This is done using a set of rules which generate a list of EECs for each character and selects default degree values for all EEC variables. The mapping from dialogue acts to EECs

concerns speaker and addressee only. However, in some cases other characters may be affected, too, for instance, because of their special *role*. In CrossTalk, Cyberella has the role of the director and is responsible for the other characters' actions. Consider the example, that Ritchie performs a verbal attack on Tina, while a user is watching the performance. The dialogue act [=attack T] creates the following EECs:

Tina: D: -1.0 (very undesirable), P: -1.0 (very blameworthy)
Ritchie: D: +0.5 (moderately desirable)
Cyberella: D: -0.5 (moderately undesirable), P: -0.5 (moderately blameworthy)

Ritchie's dialogue act is appraised by Cyberella as being undesirable, because she is responsible for the performance and does not want it to be disturbed by "internal" quarrels. For the same reason she considers it a blameworthy action by Ritchie.

Table 2: Dialogue act tags in the CrossTalk character script.

Dialogue act	Description
admire	Speaker expresses admiration for the addressee.
attack	Speaker attacks addressee by a verbal argument.
bad_joke	Speaker makes a bad joke that targets the addressee.
boast	Speaker praises him-/herself in front of the addressee.
calm	Speaker calms addressee.
chide	Speaker seriously chides addressee either for doing something wrong or doing something morally bad.
command	Speaker commands the addressee to perform some action.
congratulate	Speaker sincerely congratulates the addressee (if meant ironically, use mock).
console	Speaker consoles addressee who faces some bad event.
correct	Speaker corrects a mistake the addressee has made.
criticize	Speaker criticizes the addressee for performing an action not well enough.
doubt	Speaker expresses doubt concerning something the addressee said or did.
defend	Speaker defends his/her argument against the addressee (reaction to "attack" or "correct").
encourage	Speaker encourages the addressee to do or believe in something.
excuse	Speaker excuses her-/himself in front of the addressee.
good_joke	Speaker makes a good joke with addressee as an audience.
insult	Speaker directly or indirectly insults the addressee.
mock	Speaker mocks addressee. Weaker than an insult.
praise	Speaker praises addressee for some action or attitude.
regret	Speaker regrets some action or attitude in front of addressee.
reproach	Speaker reproaches addressee for some action or opinion.
sulk	Speaker expresses her/his being filled with indignation by the addressee's action.
tease	Speaker teases addressee. This is milder than an insult or mocking.

User interactions, like the feedback visitors can give in the ON mode via the buttons *applause*, and *boo* is treated like the dialogue acts congratulate and criticise respectively. This enables the CrossTalk characters to react emotionally on user input. Thus, by giving frequent applause and boo feedback the visitor can significantly influence the emotional state of the characters and elicit emotional responses.

6. Computing Emotions

Using dialogue act tags and appraisal tags allow the generation of EECs. This section shows how the EmotionEngine maps EECs to emotional states based on the OCC model. Our approach augments the OCC model by integrating the Five Factor Model (FFM) of personality [20] into the emotion computation process. The FFM is a descriptive model, with the five dimensions (extraversion, agreeableness, conscientiousness, neuroticism, and openness) being derived from a factor analysis of a large number of self- and peer reports on personality-relevant adjectives. In our system we concentrate on the traits extraversion, agreeableness and neuroticism because their impact on the emotional intensity development appears to be most

obvious. The traits can be individually defined for each character (see Figure 1). They affect the intensity of emotional states, i.e. a happy character tends to be more happy if the character's personality is extravert and agreeable. This is realised by biasing the emotions with a baseline intensity according to the personality setting, 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.

Emotions usually do not last forever. For the simulation of emotion decay, the EmotionEngine provides three different decay functions (see Figure 1). Emotions, however, should not a priori influence the characters behaviour measurably. For this reason, only the most dominant emotion whose computed intensity exceeds a pre-defined threshold does have an impact on the character's behaviour. For tracking changes in the emotions an emotion monitor for each character is provided as depicted in Figure 1 on the left side. A red marker at the right side of an emotion signals that this emotion influences the character's behaviour.

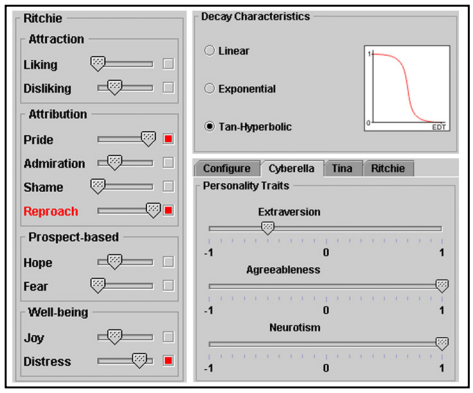

Figure1: Emotion monitor and configuration panels.

Table 3: OCC-Emotions in CrossTalk.

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.

The emotions that people experience are very much dependant on the situation they are in. The EmotionEngine supports all 22 emotion types specified by the OCC model, however only a limited number of emotions are relevant for the current CrossTalk scenario (see Table 3).

7. Expressing Emotions

We perform the expression of emotions in three respects: speech, non-verbal behaviour and handling of user interactions.

7.1 Speech

In CrossTalk we use the IBM ViaVoice TTS technology for generating speech output [21]. 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 as shown in Table 4. The table gives an qualitative overview on the parameter changes for English female and male voices. The plus letter (+) stands for an increase 20%, respective the a minus letter (-) stands for a decrease of 20%. The arrows denote a dynamic increase (or decrease) up to a maximum (minimum) value over the time the emotion is active. We compared these relationships with the literature and found them confirmed in cases where the same emotions were investigated [22].

Table 4: Emotional speech parameters.

Emotion	Tendencies		
	Pitch	Speed	Vol.
Joy, Hope	++		+
Distress, Fear	+++	+	
Admiration, Liking	++	---	+
Reproach, Disliking	--	--	+ ↑
Pride	+	--	+
Shame	+	-	-- ↓

7.2 Nonverbal Behaviour

In the first versions of CrossTalk the non-verbal behaviour of the characters was fully specified. A large repertoire of actions, about 35 actions per character, support the author in defining non-verbal behaviour. The possible actions come in four categories: gestures (G), facial expressions (F), posture shifts (P) and actions (A). Gestures were taken from a catalogue derived from analysing a German TV show with manual gesture annotation [23, 24]. Due to the use of Microsoft Agent technology for the character player a fine-grained synchronization of gesture and speech is not possible.
The appraisal and dialogue acts tags allow the automatic generation of some kinds of non-verbal behaviour. This is done by mapping dialogue acts and/or emotions to actions that are then performed by the characters:

- *Gaze*: The run-time analysis of the current speaker and of the addresses enables the system to create gaze actions for the character, e.g. if the visitor gives feedback during ON mode, all character may give the user a glance signalling that his/her feedback is recognized.
- *Conversational gestures*: In some cases gestures are triggered to reflect a characters dialogue act, e.g. the emblematic gesture of a raised forefinger for the dialogue act *correct*. In the case that Ritchie attacks Tina, he will automatically use a pointing gesture at Tina, when he utters the attack.
- *Facial expressions*: Facial expressions are used to signal positive or negative emotions.

In order not to lose control over the characters' behaviour, the script writer can always override the use of the automatic triggered gestures, by specifying a gesture at this position in the script.

7.3 User Interaction

Our approach for creating presentations is based on the separation of scene flow and scene content. While the speech and non-verbal behaviour is specified in the context of a scene, the user interaction is specified through the transitions between scenes. In CrossTalk authors have several possibilities to influence the scene transitions (see [9,

17]). With their help, authors can specify transitions according to emotional states in order to show emotional reaction in behaviour.
An example is the treatment of the visitor's feedback during the car sales dialogue in ON mode. There are three types of feedback: applause, boo, and question. The feedbacks of applause and boo have some impact on the characters emotional state. As a reaction to the feedback the characters interrupt the car sales dialogue and comment on the feedback. This is done by selecting an appropriate scene. The selection of this scene does not only depend on feedback but also on the characters' emotional state. In the case that the visitor gives only negative feedback (boo) for a certain period of time, Cyberella becomes more and more *distressed.* This results in the transition to a special scene: Cyberella addresses the visitor's being displeased and therefore suggests to change some parameters and rerun the show.

8. Conclusions and Future Work

The use of the EmotionEngine for computing emotions in the CrossTalk framework driven by the extensions of the dialogue scripting language allows the generation of emotional expression in character behaviour. Together with context knowledge, provided by the ContextMemory module, it enables the characters to adapt to user feedback and to react to previous encounters with users in an emotional way.
In contrast to the often used direct mapping of world aspects (events, actions, and objects) to emotional expression, the use of the EmotionEngine allows the combination of more than one world aspect which can be used for a fine-grained generation of emotional expressions. The emotion decay also helps to create a believable emotional behaviour by fading emotional expressions according to the decaying emotions. Moreover, this approach prepares the system for using emotions in the automatic, plan-based generation of dialogues. We will investigate how emotions affect the selection of dialogue strategies.

Acknowledgements

This work has been built upon contributions from the VirtualHuman project funded by the German Ministry for Education and Research and from the EU-funded IST project NECA.

References

[1] Hayes-Roth, B., van Gent, R.: Story-Making with Improvisational Puppets. In: Proc. of Autonomous Agents'97, 1997, 92-112
[2] 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 Autonous Agents'01, 2001, 409-416

[3] Laurel, B. Computers as theatre. Addison-Wesley, Reading MA, 1993
[4] Murray, J.H. Hamlet on the holodeck: The future of narrative in cyberspace. The MIT Press, Cambridge MA, 2000
[5] 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
[6] Prendinger H., Ishizuka M. SCREAM: Scripting emotion-based agent minds. In: Proc. of AAMAS'02. ACM Press. 2002; 350-351
[7] Rist, T., and 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, 79-84
[8] André E., Rist T. Presenting through performing: On the use of multiple animated characters in knowledge-based presentation systems. In: Proc. of IUI'00. ACM Press. 2000, 1-8
[9] Patrick Gebhard, Michael Kipp, Martin Klesen, Thomas Rist. Authoring Scenes for Adaptive, Interactive Performances, In: Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-03), 2003. (in press)
[10] Joseph Bates. The role of emotion in believable agents. Communications of the ACM, 37(7):122–125, 1994
[11] Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill, editors. Embodied Conversational Agents. The MIT Press, Cambridge, Massachusetts, 2000
[12] Ana Paiva, editor. Affective interactions: towards a new generation of computer interfaces, volume 1814 of Lecture Notes in Artificial Intelligence. Springer-Verlag, Berlin, Heidelberg, New York, 2000
[13] Robert Trappl and Paolo Petta, editors. Creating Personalities for Synthetic Actors: Towards Autonomous Personality Agents, volume 1195 of Lecture Notes in Computer Science. Springer-Verlag, Berlin, Heidelberg, New York, 1997
[14] Ishizuka M., Tsutsui T., Saeyor S., Dohi H., Zong Y., Prendinger H. MPML: A multimodal presentation markup language with character control functions. In: Proc. of the Workshop on Achieving Human-like Behavior in Interactive Animated Agents held in conjunction with the Fourth International Conference on Autonomous Agents, 2000, pp. 50-54
[15] Marriott, A., Stallo, J., VHML Uncertainties and problems. A discussion. In: Proceedings AAMAS-02 Workshop on Embodied conversational agents – let's specify and evaluate them!, 2002
[16] De Carolis, B., Pelachaud, C., Poggi, I., de Rosis, F.: Behavior planning for a reflexive agent. In: Proceedings 17th International Conference on Artifcial Intelligence (IJCAI-01), 2001, pp. 1059-1066
[17] Martin Klesen, Michael Kipp, Patrick Gebhard, Thomas Rist, 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 (in press)
[18] Microsoft® Agent is a software technology: http://www.microsoft.com/msagent/, 2003
[19] Andrew Ortony, Gerald L. Clore, and Allan Collins. The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, MA, 1988.
[20] R.R. McCrae and O.P. John. An introduction to the five-factor model and its implications. Journal of Personality, 60:171–215, 1992.
[21] IBM ViaVoice TTS technology: http://www-3.ibm.com/software/speech/dev/, 2003
[22] M. Schröder (2001). Emotional Speech Synthesis - A Review. Proc. Eurospeech 2001, Aalborg, Vol. 1, pp. 561-564
[23] Kipp M. Anvil – a generic annotation tool for multimodal dialogue. In: Proc. of the 7th European Conference on Speech Communication and Technology (Eurospeech). 2001; 1367-1370
[24] Kipp M. Analyzing individual nonverbal behavior for synthetic character animation. In: Oralité et Gestualité - Actes du colloque ORAGE. Cavé C., Guaitella I., Santi S. eds. 2001