# The NECA Project: Net Environments for Embodied Emotional Conversational Agents

Brigitte Krenn
Austrian Research Institute for Artificial Intelligence ÖFAI
Vienna, Austria
brigitte@ai.univie.ac.at
http://www.ai.univie.ac.at/NECA/

## Introduction

The purpose of the NECA project (http://www.ai.univie.ac.at/NECA/) is to develop a platform for the implementation of emotional conversational agents for Web-based applications. The users watch embodied conversational agents engage in verbal and nonverbal interaction in virtual "locations" on the internet. In the demonstrators, eShowroom and Socialite, two such "locations" are realised.

While in the first project year the core techniques, programs and resources to enable the generation of animated conversation in web-based applications have been developed and implemented, the second project year is dedicated to the refinement of the system modules and the enhancement of the functionality of the demonstrators, with a focus on affective reasoning, and verbal and nonverbal expression of personality and affect in a situation-specific and socio-culturally mediated way.

In the following we give an introduction to the demonstrators, followed by an overview of the NECA architecture, and then describe strategies for emotion modelling in NECA.

## The Demonstrators

In the **eShowroom** scenario a car sales dialogue between a seller and one or more buyers is simulated. The purpose of this application is to entertain the site visitor and to embed product information into a narrative context similar to TV commercials nowadays. User interaction is restricted to setting general parameters prior to the display. These parameters include the user's preferences in respect to different value dimensions, e.g. on how important aspects like sportiness, prestige or environmental issues are for the user. After having specified these preferences, a scene is generated which takes these settings into account: The agents/interlocutors will put special emphasis on conveying information about those aspects, which have been classified as being of importance for the user. The user can also specify the personality traits of the agents, i.e., their agreeableness and politeness, to influence the style and the course of their conversation. This option aims to provide a means of entertaining the user by experimenting with different (possibly absurd) settings.

The **Socialite** demonstrator implements a multi user web-application in the social domain. The users create their personal avatar, endow it with personality traits and preferences and send it to the virtual environment in order to meet other avatars. The overall goal is to be accepted in the community, to

reach a certain degree of popularity within this environment. In this setting the user is not permanently logged on. The avatar/agent will report back to the user about encounters with other avatars when the user logs in the next time. This report is presented in the form of monologues, which are alternated with displays of dialogues between avatars, much in the style of the rendering of retrospectives in older movies. The user is then queried for choosing new instructions for the avatar from a given set of possibilities and sends the avatar off to its environment again.

The demonstrators can be accessed via the NECA homepage www.ai.univie.ac.at/NECA/.

## NECA Architecture and Rich Representation Language

The NECA architecture consists of the following main components: a scene generator, a multimodal natural language generator, a text/concept-to-speech synthesis, and a gesture assignment module. The information available at the interfaces between the system components is encoded in an abstract XML-compliant representation scheme which all together constitutes the NECA RRL (Rich Representation Language). See Figure 1. The NECA architecture is described in more detail in [Krenn et al., 2002]. For a description of the RRL see [Piwek et al., 2002].

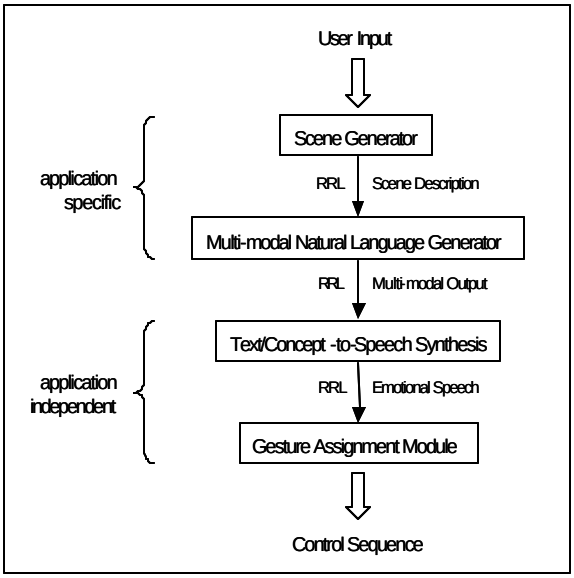


**Figure 1: Overview of the NECA Architecture**

The *Rich Representation Language* has been designed for the description of agent behaviour in our net environments. The RRL represents a wide range of expert knowledge required at the interfaces between the different components in the NECA architecture. The RRL differs from other multimedia markup languages in that these are typically designed to support a fairly text-based annotation of multimodal input to media players, ideally in a rather generalised and standardised way, whereas the RRL is in addition capable of representing expert knowledge which may be created by a processing component rather than a human author. In developing the RRL we draw as much as possible on existing standardisation efforts and build on well-defined cores of XML-based markup languages. In

addition, particular care is taken in developing a mechanism that allows for a fine-tuned integration of verbal and nonverbal aspects of communication.

In the preparation phase of an application the user is able to provide the system with information on her information needs and preferences ("User Input"). In the application scenarios the *scene generator* takes the role of a playwright, generating a script for the characters that become actors in a scene. In the script, the communicative acts to be carried out are specified as well as their temporal coordination and the emotion associated.

The *multimodal natural language generator* transforms the formal specification of the communicative acts into text, annotated with syntactic, semantic, and pragmatic features. The component is also responsible for selecting meaning-bearing and (discourse) functional 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 synchronisation of verbal and nonverbal aspects of behaviour in the *gesture assignment module*. The gesture assignment module also schedules physiologically based animations (e.g. eye blinking and breathing) in accordance with the constraints imposed by the content-bearing gestures, so as to make the characters more life-like. The output of this process is a *control sequence* comprising the synchronised verbal and nonverbal behaviour of all the characters in the scene. In a last step this control sequence is converted into a data stream that can be processed by a specific player, such as Macromedia Flash or Microsoft Agent.

## Aspects of Emotion Modelling in NECA

Within the NECA project we make use of two approaches to emotion modelling. One is based on the OCC model [Ortony, Clore, Collins, 1988] and is applied for specifying and reasoning with emotion types which are assigned to communicative acts. The other approach is based on emotion dimensions (see for instance Cowie et al., 1999) and is used in speech synthesis in order to express shades of emotions in the voice quality and thus allows to convey changes of emotional tone over time.

With regard to emotion types assigned to communicative acts, we distinguish between the emotions which the speaker feels when performing the act and the ones which s/he expresses in performing the act. Additionally, both the emotion felt by the hearer when the act is performed and the emotions expressed by her or him as a direct result of the dialogue act are part of the description of a communicative act. The emotion assigned influences the realization of the utterance -- for instance the lexical selection of adjectives --, as well as the generation of gestures and facial expressions accompanying the act. Emotional state is also reflected in physiologically motivated characteristics of the animation such as the intensity of eye blinks and breathing, which is modelled in the gesture assignment module.

In speech synthesis, we introduce an innovative approach to making synthesis flexible and expressive, by recording diphone voices with three phonetically defined voice qualities, and by explicitly modelling the acoustic properties of emotions described in terms of emotion dimensions, see [Schröder et al., 2001] and [Schröder, 2003]. The rich segmental material, in combination with the flexible control of prosodic features in diphone synthesis technology, enables us to implement a

mapping from emotion dimensions to acoustic properties of the synthetic voice. This results in unprecedented flexibility in emotion expression through synthetic speech.

Another innovation of the NECA project is that a mapping between the OCC emotion types and the emotion dimensions is developed.

## Acknowledgements

The Austrian Research Institute for Artificial Intelligence (ÖFAI) is supported by the Austrian Federal Ministry for Education, Science and Culture. The work described here is joint work of the members of the NECA consortium. This research is supported by the EC Project NECA IST-2000-28580. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

## Bibliography

[Cowie et al., 1999] Cowie, R., Douglas-Cowie, E., Appolloni, B., Taylor, J., Romano, A., Fellenz, W. (1999). In Mastorakis, N. (ed.). Computational Intelligence and Applications. World Scientific & Engineering Society Press.

[Krenn et al., 2002] Krenn, B., Grice, M., Piwek, P., Schröder, M., Klesen, M., Baumann, S., Pirker, H., van Deemter, K., Gstrein, E. (2002). Generation of Multi-modal Dialogue for Net Environment. In *Proceedings of KONVENS-02*, 30 September - 2 October 2002, Saarbrücken, Germany.

[Ortony, Clore, Collins, 1988] Ortony, A., Clore, G. and Collins, A. (1988). The Structure of Emotions. Cambridge University Press. Cambridge MA.

[Piwek et al., 2002] Piwek, P., Krenn, B., Schröder, M., Grice, M., Baumann, S., Pirker, H. (2002). RRL: A Rich Representation Language for the Description of Agent Behaviour in NECA. In *Proceedings of the Workshop ``Embodied conversational agents - let's specify and evaluate them!''*, held in conjunction with AAMAS-02, July 16 2002, Bologna, Italy.

[Schröder et al., 2001] Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M. and Gielen, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. In *Proceedings of Eurospeech*, Aalborg, Denmark. Volume 1, 87-90.

[Schröder, 2003] Schröder, M. (2003). Experimental study of affect bursts. In *Speech Communication Special Issue Speech and Emotion*, 2003.