Representational Lego for ECAs

Brigitte Krenn, OFAI

A proposal for contribution to the HUMAINE WP6 Workshop on "Emotion and Interaction"
Topics: - representation of affective multimodal behaviours
- flexibility and interoperability of representations
- pluggability of system components

Abstract

This contribution is an appeal for a consequent use of XML representation languages for the representation of ECA relevant core concepts, and for use at the interfaces of the system components constituting an ECA system. A list of core aspects constituting a fully powered ECA is presented which is meant to be a basis for a broad discussion of core components and related concepts central to the development of interactive affective ECAs. The hope is that at the end of the workshop the inventory will be further refined.

Finally the following immediate steps within HUMAINE WP6 "Representation Languages" are proposed: examine existing representation and scripting languages for a common core by identify common concepts and by comparing their representations; relate these concepts to system architectures for interactive affective ECAs and their components; develop unified XML representations for these core concepts; make different representations interoperable.

1 Introduction

From practical experience on modeling and representing high- and low-level concepts relevant at different stages of processing in ECA systems, we have learned that it is an illusion to define a comprehensive representational standard that is widely reused in the community. In fact, everybody designs their own representations with the risk of constant reinvention of the "representational" wheel.

Up to now, a variety of systems featuring ECAs exist, integrating a number of subsystems such as speech synthesizers, multimodal natural language generators, affective reasoners, etc. For some of these components a variety of off-the-shelf modules are available, other components are still in their early stages of research and development. In this situation a plug and play approach to ECA systems is desirable which facilitates the integration of various, highly specialized components, and eases the replacement of components.

Moreover as available overall standards are obviously not accepted, and the definition of representations on a case to case basis is a waste of time, we argue

for exploring the middle way, i.e. to identify core concepts relevant at certain stages of processing in ECA systems, and to define high-level representations and their low-level counterparts at the basis of core concepts.

In this contribution, we argue that XML-compliant representation languages used as representation formats at the interfaces of ECA system components is a good choice, because XML representations are flexible, foster interoperability and are compatible with a wealth of processing and programming tools for the manipulation of XML documents.

Representation, markup and scripting languages are discussed in section 2. In section 3, arguments for the advantages of XML-based interfaces in ECA systems are presented. In section 4, concepts central to ECA systems are listed.

2 Representation versus markup versus scripting languages

Markup languages typically define sets of markups that give the non-expert user (usually a web designer) the possibility to annotate a text with high-level expert information. See for instance VoiceXML[1] for creating voice enabled web applications, or VHML[2] (Virtual Human Markup Language) for creating interactive multimodal applications with Talking Heads or full bodied ECAs. Other examples of ECA markup languages where text is annotated with high-level concepts are APML (De Carolis et al. 2002) or MPML (Zong et al. 2000).

Representation languages on the contrary allow for technically detailed annotations of theory-specific information. Thus representation languages are well suited to function as data representation formats inside a system, especially as representations at the interfaces between the individual subcomponents. The NECA RRL[3] (Rich Representation Language, Piwek et al., 2002), for instance, defines XML documents at the in- and output of system components. The components of the NECA platform for generating simulated animated affective dialogue (Krenn 2003) operate directly on XML documents by interpreting (input) and completing (output) information.

While markup and representation languages cover the declarative part, **scripting languages** also incorporate procedural knowledge such as if-than-else or do statements. Thus scripting languages are comparable to high-level programming languages. Examples in the field of ECAs are STEP and its XML variant XSTEP (Huang et al. 2003), and ABL (Mateas & Stern 2004).

On the one hand, markup languages are indispensable in application development, because the application designer needn't be an expert in the fields underlying the development of ECA systems. On the other hand, representation languages are crucial in research contexts, because of the necessity to represent highly specific, low-level information. All in all, the representations languages we aim at must, for practical reasons, provide both high- and low-level representations, and also mechanism or guidelines to translate between

[1] http://www.voicexml.org
[2] http://www.vhml.org
[3] http://www.oefai.at/NECA/RRL/

high-level concepts and their low-level realizations. For theoretical clarity and to ease extendibility the declarative and the procedural aspects must be clearly separated.

3 Advantages of XML interfaces in ECA systems

Using XML-based representation languages at the interfaces of modules in ECA systems has a number of advantages.

Information encoding: XML is a flexible and easy to share means for providing highly standardized representation frameworks, and thus is particularly suitable for encoding and storing the different kinds of information required at the input and available at the output of the modules within an ECA system.

Integration and replacement of system components: The development of ECA systems is a complex endeavor for which it is inappropriate to build systems from scratch over and over again, a common practice up to now. The use of XML-based representations for encoding the information flow between system components opens up promising ways to reuse and easily exchange components with minimal recoding effort, for instance by employing XSL Transformations (XSLT) or by using the Document Object Model (DOM) programming API. An example for the use of XSLT stylesheet technology for module integration between two XML-based TTS systems is described in (Schröder & Breuer, 2004). Reducing the integration costs is crucial to bundle efforts and advance ECA technology.

Developing mock-up systems: The use of XML documents allows for a script based approach through exporting and importing XML-based interface representations to and from textual format, which enables the system developer to manipulate the in- and output of selected modules by hand. Scripting of XML interface documents is a practical means for developing test versions of ECA systems, as it offers a broad range of possibilities for manual intervention either by piecemeal modification of XML documents generated from the output of individual system components or by specifying the anticipated output of not yet implemented system components.

4 Representation languages: format requirements, ECA relevant components and concepts

A number of languages for markup, representation or scripting of multimodal (affective) agent behaviour exist. As regards their content, approaches overlap in some core areas as well as differ depending on their application domain and on implementation specific aspects. Currently available representation and scripting languages for ECAs are largely designed for multimodal presentation. Some work is more related to multimodal dialogue (e.g. APML, RRL), other work is more focused on motion, e.g. XSTEP, MURML (Kransted et al. 2002).

Format requirements: What we need are languages that combine high-level markups and detailed representations depicting information relevant at certain stages of processing. The languages need to be extendable as regards the

phenomena described (width) and the granularity of the descriptions (depth). The individual descriptions need to be flexibly combinable, and existing XML-based specifications must be easily embeddable.

Inventory of information relevant for ECA systems: The following is a list of core aspects constituting a fully powered ECA. The list is far from being exhaustive, none the less it is meant to be a basis for a broad discussion of core components and related concepts central to the development of interactive affective ECAs. The hope is that at the end of the workshop the inventory will be further refined.

World parameters: comprising a definition of the layout of the overall scenery, the objects and characters (ECAs and user characters) present; e.g. RRL (character specifications including character name, appearance, personality and role, selected voice, speech rate and pitch) , PML (RRL specifications extended by specifications for a 3D environment including virtual and user characters, and objects; Klesen & Gebhard 2004).

Scenes and story lines: such as representations of the common grounds, dialogue history and displayed behaviours. As regards the representation of common ground we can learn from natural language generation, for modeling the dialogue history <seq> and <par> elements as designed in SMIL[4] are employed, as regards behaviours and story lines one can take advantage of the work on interactive drama realized in Facade (Mateas & Stern 2002).

Dialogue: dialogue relevant information comprises labels for speech or communication acts, and labels representing information structure. As regards speech or communication acts one can resort to a body of general acts as defined in DAMSL[5], but from practical work we know that each application requires its very specific dialogue acts too.

Speech: core representations comprise SAMPA for phonetic transcription, TOBI labels for representing prosodic information; e.g. the RRL provides an elaborate inventory of speech related information.

Body: high-and low-level body orientation and direction references, e.g. XSTEP or MURML; low-level body representations, e.g. H-Anim[6] joints; it is also important to be able to define macros for complex (recurrent) movements, see for instance the <action> element in XSTEP.

Face: facial expression is a major means to convey emotional state; high-level representations typically employ emotion labels such as happy, sad etc.; MPEG4 FAPs define low-level representations for facial expression;[7]

What is also needed are reusable libraries for facial expression, gesture, and motion, which establish a principled relation between high- and low-level descriptions. See for instance (Hartmann et al. 2002) and (Krenn & Pirker 2004) for initial proposals of gesture repositories.

Personality and emotion: in the ECA community two approaches to personality and emotion are widely employed, the Five-Factor (OCEAN) model

[4]http://www.w3.org/AudioVideo/

[5]http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/

[6]http://www.h-anim.org/

[7]A complete table of MPEG4 FAPs see on
http://www-dsp.com.dist.unige.it/ pok/RESEARCH/MPEG/fapspec.htm.

(Wiggins 1996) and the OCC model (Ortony et al. 1988), respectively. This is reflected in the inventory of several representation languages for ECAs, e.g. PAR (Allbeck & Badler 2002) makes provisions for OCC and OCEAN; MPML V2.0e incorporates labels for the 22 emotions defined in OCC (Zong et al. 2000); RRL encodes the OCC labels plus their extension by Elliot (1992), a subset of OCEAN labels and a politeness attribute; APML defines its own set of emotion labels geared towards the communication situation (medical counselling) and the type of ECA (a talking head) used.

While OCC is a cognitive model of appraisals, emotion correlates in speech are typically modeled by means of a dimensional approach (Schröder 2004), and emotions expressed via the face are typically modeled by means of Ekman's emotion categories (Ekman 1993).

Especially in the context of HUMAINE we expect new emotion theories and cognitive models to be developed. These will also lead to new implementations of system components. Again there will be a need for the incorporation of these novel insights into the representations, as well as a requirement for an easy exchange of components.

Temporal control and synchronization: in the ECA domain, synchronization is required for (i) the multi-modal behaviour of individual agents, and (ii) the temporal ordering of actions/behaviours of agents interacting with the outside world, i.e. agent-to-agent interaction, agent-to-user interaction and agent-object interaction. As regards the multimodal behaviour of an agent, speech is the guiding medium with phoneme durations (in milliseconds) as smallest units. With the availability of fine-grained prosodic information, facial and gesture animation can be time-aligned to individual phonemes, accented syllables or boundaries of intonation phrases. From a motion oriented point of view, however, beats are proposed as smallest units. See for instance XSTEP where time reference is given in beats (1 beat = 1 second) and tempo (= number of beats per minute). Different speeds can be modeled on the basis of beat durations. What is still missing, is an integrated approach where not only speech defines the timing of its accompanying facial expressions and gestures, but also motor activation constrains voice quality. Accordingly the representations for temporal control and synchronization need to be rather flexible to be able to interoperate with the different possibilities to model behaviour timing.

In ECA representation languages, typically <par> and <seq> tags are used which are in the best case interoperable with the SMIL timing and synchronization module[8]. In addition to the SMIL operators XSTEP defines operators influenced from dynamic logic, such as non-deterministic choice, repeat, execution (= make statement true), and conditional.

Interactivity: As true interactivity becomes more and more a topic in ECA systems – see for instance the German Virtual Human project[9] or Facade (Mateas & Stern 2004) – a further strand of expansion of ECA representation languages is opened up. Some of the current major questions are: What are the desired smallest communicative units, from a point of view of speech, dialogue,

[8] http://www.w3.org/TR/SMIL2/smil-timing.html
[9] http://www.virtual-human.org/

and interactive drama? What are the technically manageable "smallest" units? What are the challenges to the technological state-of-the art of individual system components.

5 Future Steps

Future steps are manifold, such as

to define an inventory of core concepts and their high- and low-level representations in XML;

to integrate dialogue oriented and motion oriented approaches;

to incorporate perception (interpreted sensory input);

to make representations interoperable;

to model interactivity;

to devise reusable libraries for gesture, facial expression, and motion.

Immediate steps within HUMAINE WP6 "Representation Languages": examine existing representation and scripting languages for a common core by identify common concepts and by comparing their representations; relate these concepts to system architectures for interactive affective ECAs and their components; develop unified XML representations for these core concepts; make different representations interoperable.

Make publicly available via the Internet a body of core XML representations, encourage the development of improved, complementary representations interoperable with the core to be collected at the web site for common use.

6 Literature

Allbeck J., Badler N. Towards Representing Agent Behaviors Modified by Personality and Emotion, in Marriott A. et al. (eds.), Embodied Conversational Agents: Let's Specify and Compare Them!, Workshop Notes, Autonomous Agents & Multiagent Sytems 2002, University of Bologna, Bologna, Italy, 2002.

De Carolis B., Carofiglio V., Bilvi M., Pelachaud C. APML, a Markup Language for Believable Behavior Generation in Marriott A. et al. (eds.), Embodied Conversational Agents: Let's Specify and Compare Them!, Workshop Notes, Autonomous Agents & Multiagent Sytems 2002, University of Bologna, Bologna, Italy, 2002.

P. Ekman. Facial expression of emotion. In American Psychologist, 48:384–392, 1993.

Elliott C.D. The Affective Reasoner: A process model of emotions in a multi-agent system, Northwestern University, Illinois, Ph.D. Thesis, 1992.

Hartmann B., Mancini M., Pelachaud C. Formational parameters and adaptive prototype instantiation for MPEG-4 compliant gesture synthesis,Computer Animation, Geneva, June 2002.

Huang Z., Eliens A., Visser C. XSTEP: a Markup Language for Embodied Agents, Proceedings of the 16th International Conference on Computer Animation and Social Agents (CASA'2003), IEEE Press, 2003.

Klesen M., Gebhard P. Player Markup Language. Version 1.2.4, DFKI, internal document. March 2004.

Kransted A., Kopp S., Wachsmuth I. MURML: A Multimodal Utterance Representation Markup Language for Conversational Agents, in Marriott A. et al. (eds.), Embodied Conversational Agents: Let's Specify and Compare Them!, Workshop Notes, Autonomous Agents & Multiagent Sytems 2002, University of Bologna, Bologna, Italy, 2002.

Krenn B. The NECA Project: Net Environments for Embodied Emotional Conversational Agents Project Note, KI - Künstliche Intelligenz Themenheft Embodied Conversational Agents, Springer-Verlag, 2003.

Krenn B., Pirker H. Defining the Gesticon: Language and Gesture Coordination for Interacting Embodied Agents, in Proceedings of the AISB-2004 Symposium on Language, Speech and Gesture for Expressive Characters, University of Leeds, UK, pp.107-115, 2004.

Mateas M. Stern A. Architecture, Authorial Idioms and Early Observations of the Interactive Drama Facade Technical Report CMU-CS-02-198, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA. December 2002.

Mateas M., Stern A. A Behavior Language: Joint Action and Behavioral Idioms. In Prendinger H., Ishizuka M. (eds). Life-like Characters. Tools, Affective Functions and Applications. Springer, 2004

Ortony A., Clore G.L., Collins A. The Cognitive Structure of Emotions, Cambridge University Press, Cambridge, UK, 1988.

Piwek P., Krenn B., Schroeder M., Grice M., Baumann S., Pirker H. RRL: A Rich Representation Language for the Description of Agent Behaviour in NECA, in Marriott A. et al. (eds.), Embodied Conversational Agents: Let's Specify and Compare Them!, Workshop Notes, Autonomous Agents & Multiagent Sytems 2002, University of Bologna, Bologna, Italy, 2002.

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. 2004.

Schröder M., Breuer S. XML Representation Languages as a Way of Interconnecting TTS Modules, Proceedings of ISCLP'04. Jeju, Korea., 2004.

Wiggins J. The Five-Factor Model of Personality: Theoretical Perspectives. The Guilford Press, New York, 1996.

Zong Y., Dohi H., Ishizuka M. Multimodal Presentation Markup Language MPML with Emotion Expression Functions Attached. Proceedings of the International Symposion on Multimedia Software Engineering (IEEE Computer Soc.), Taipei, Taiwan 2000.