Representing Emotions and Related States in Technological Systems

Marc Schröder and Hannes Pirker and Myriam Lamolle and Felix Burkhardt and Christian Peter and and Enrico Zovato

Abstract In many cases when technological systems are to operate on emotions and related states, they need to represent these states. Existing representations are limited to application-specific solutions that fall short of representing the full range of concepts that have been identified as relevant in the scientific literature. The present chapter presents a broad conceptual view on the possibility to create a generic representation of emotions that can be used in many contexts and for many purposes. Potential use cases and resulting requirements are identified and compared to the scientific literature on emotions. Options for the practical realisation of an Emotion Markup Language are discussed in the light of the requirement to extend the language to different emotion concepts and vocabularies, and ontologies are investigated as a means to provide limited "mapping" mechanisms between different emotion representations.

Marc Schröder
Deutsches Forschungszentrum fr Knstliche Intelligenz, Saarbrücken e-mail: schroed@dfki.de

Hannes Pirker
Austrian Research Institute for Artificial Intelligence, Austria e-mail: hannes.pirker@ofai.at

Myriam Lamolle
Université Paris VIII, France e-mail: m.lamolle@iut.univ-paris8.fr

Felix Burkhardt
Deutsche Telekom Laboratories, Germany e-mail: felix.burkhardt@t-systems.com

Christian Peter
Fraunhofer IGD, Germany e-mail: christian.peter@igd-r.fraunhofer.de

Enrico Zovato
Loquendo S.p.A., Italy e-mail: enrico.zovato@loquendo.com

1 Aims and purpose

Machines that register when a user is emotional, machines that express emotions, and machines that reason about the appropriate emotion in a given situation - what used to be regarded as science fiction not long ago is starting to become a reality. Many systems that implement one or the other aspect of this kind of behaviour are being built and show first successes in individual aspects of these complex tasks. From the point of view of applications, the modelling of emotion-related states in technological systems can be important for two reasons.

1. To enhance computer-mediated or human-machine communication. Emotions are a basic part of human communication and should therefore be taken into account, e.g. in emotional chat systems or emphatic voice boxes. This involves specification, analysis and display of emotion related states.
2. To enhance systems' processing efficiency. Emotion and intelligence are strongly interconnected. The modelling of human emotions in computer processing can help to build more efficient systems, e.g. using emotional models for time-critical decision enforcement.

For early systems, it is acceptable - even inevitable - to use ad hoc representations. For example, an early emotion recognition system may use simple words such as "happy", "angry", "sad" to describe its recognition outputs; an emotion-related reasoner program may conclude that a "good event" should trigger the emotion of "joy"; and an early expressive system may mix up behavioural with emotional labels, and use "happy" or "smiling" interchangeably. Such ad hoc solutions, tailor-made to the requirements of the immediate application domain, are fine as long as the various systems are not trying to communicate with each other.

However, more recently, integrated emotion-oriented computing systems are appearing, which typically consist of multiple components covering various aspects of data interpretation, reasoning, and behaviour generation. In this case, the need for a standardised way of representing emotions and related states is becoming clear: emotion-related information needs to be represented at the interfaces between system components in a way that allows one system component to make sense of the output from another component. We call such a representation an "emotion markup language". This chapter describes a number of basic considerations that should be addressed by an emotion markup language.

A standardised way to mark up the data needed by such "emotion-oriented systems" has the potential to boost development both in academia and industry primarily because

(a) data that was annotated in a standardised way can be interchanged between systems more easily, thereby simplifying the reuse of emotional databases;
(b) the standard can be used to facilitate the creation of reusable submodules of emotion processing systems, e.g. submodules for the recognition of emotion from text, speech or multimodal input;

(c) the use of a standardised technology in a new endeavor can help to ensure that best-practice experiences from previous projects are taken into account.

The present contribution provides an overview of work in the area of representing emotions in technological systems. A number of existing markup and representation languages containing elements of emotion representation are briefly presented. However, the main focus of the chapter is on the considerations that should underlie a standardised, reusable representation. In that context, the chapter draws on recent work in the W3C Emotion Incubator Group (`http://www.w3.org/2005/Incubator/emotion/`), an international endeavour investigating the prospects of defining a general-purpose Emotion annotation and representation language. We present the outcomes of the group's work regarding the kinds of application scenarios ("use cases") which would benefit from a standardised emotion representation, and the requirements towards such a representation that arise from the use cases. We discuss how these requirements relate to existing scientific work in the area of emotion research. Finally, we formulate a number of alternatives for representing the emotions on a technical level, pointing out strengths and weaknesses of various design choices.

2 Related work

Wherever emotion-related behaviour is to be analysed from user behaviour or generated in system behaviour, there is a need to represent emotional states and related information. Thorough and scientifically well-founded representations are being proposed in the context of data annotation, and are described elsewhere in this book (Douglas et al. 2090).

Representations aimed at being used in technological systems, on the other hand, have generally been shaped by application concerns; indeed, the investigation of appropriate means and models to represent emotion related states in technological systems is going on since first such trials have been performed. In emotion recognition research, the preliminary state of the art - only a small number of distinct states can be reasonably recognised - has required only simple class labels to represent the emotional states. In research systems generating emotional behaviour, on the other hand, emotion representations have been built into several markup languages. For example, the Virtual Human Markup Language VHML (Gustavsson et al. 2001) was created in order to control the behaviour of animated characters (virtual humans); in addition to markup for facial animation, speech synthesis, dialogue management etc., the specification also contains a section for representing emotions. The actual representations are very simple: a set of nine emotions is encoded directly as XML elements, e.g.:

Example 1. Representation of a simple emotion

<afraid intensity="40">

Do I have to go to the dentist?
</afraid>

The Affective Presentation Markup Language APML (de Carolis et al. 2004) provides an attribute "affect" to encode an emotion category for an utterance (a "performative") or for a part of it:

Example 2. Affective Presentation Markup Language (APML)
<performative affect= "afraid">
Do I have to go to the dentist?
</performative>

The Rich Representation Language RRL (Krenn et al. 2002) uses an element "emotion", embedded in a dialogue act, to represent the emotion. The emotion category and its intensity can be expressed, as well as the three emotion dimensions "activation", "evaluation" and "power". In addition, there is a conceptual distinction between feeling and expressing an emotion:

Example 3. Rich Representation Language (RRL)
<dialogueAct>
...
<emotion>
<emotionExpressed type="afraid" intensity="0.3" activation="0.3"
evaluation="-0.6" power="-0.3"/>
</emotion>
<sentence><text>Do I have to go to the dentist?</text>...</sentence>
</dialogueAct>

All these languages include the representation of an emotional state as one aspect in a complex representation oriented towards the generation of behaviour for an embodied conversational agent (ECA). None of the representations aim for reusability in different contexts, and none reach a representational power coming anywhere near the complexity considered to be necessary in emotion research (see e.g. (Cowie et al. 2090)).

The Emotion Annotation and Representation Language EARL (Schröder 2006) was introduced as an attempt to address both issues: reusability and a representation approaching what is considered scientifically necessary. It can represent emotions alternatively in terms of categories, dimensions or appraisals; the intensity of the state can be indicated; several kinds of regulation are previewed, e.g. the simulation, suppression or amplified expression of an emotional state; complex emotions can be represented, as in situations of regulation or when more than one emotion is present. For example:

Example 4. Emotion Annotation and Representation Language (EARL)
<emotion category="afraid" intensity="0.4" suppress="0.6" activation="0.3"
evaluation="-0.6" power="-0.3">
Do I have to go to the dentist?
</emotion>

In view of generic use, several ways of defining the scope of the emotion were previewed in EARL: simply embedding the annotated content (as in the above example), cross-linking, or the indication of start and end times. Furthermore, the actual lists of categories, dimensions and appraisals to use were designed to be flexible, giving the user a choice of using a label set appropriate for a specific use.

One major difference between EARL and ECA-related languages such as VHML, APML or RRL is the fact that EARL explicitly excludes the description of behaviour, linguistic structures, facial expressions etc. It aims to be a specialised, plug-in language to be used in combination with other languages. The advantage of this design approach is that it is easier to add emotion representation to a variety of systems. In particular, where a system already exists, it is possible to complement the existing representations with an emotion plug-in language such as EARL. For example, a multimodal dialogue system could add some emotional competence by using EARL at the interface between its processing components: the output of a facial analysis system could include a classifier's estimation of the emotions present and the respective probabilities, encoded in EARL; a subsequent dialogue manager could interpret the emotion, and generate an emotionally coloured reaction, again using an EARL representation to transmit the intended expression to the audio and visual generation components. The use of a standardised emotion representation language would increase the chance that components can be reused and/or integrated in different systems.

The difficulty faced by any standard, and especially in the context of emotions where so many different views exist on what an emotion is and how it can be defined and described, is satisfying everyone without becoming overly complex. A language will only be used if it provides what users need; at the same time, the language should not become too complicated, so that it is easy enough to understand how to use it. For that reason, it is essential to get a clear picture of what potential users require from a markup language. The following two sections describe, in some detail, how the W3C Emotion Incubator Group has attempted to answer that question.

3 Use cases

In order to compile a new technological framework, it is good practice to start with a collection of descriptions of situations in which the need for such a framework might arise ("use cases"). So as a first step, the Emotion Incubator group gathered together as complete a set of use cases as possible for the language, with two primary goals in mind: to gain an understanding of the many possible ways in which this language could be used, including the practical needs which have to be served; and to determine the scope of the language by defining which of the use cases would be suitable for such a language and which would not. Individual use cases were grouped into three broad categories: Data Annotation, Emotion Recognition and Emotion Generation. Many of the individual use case scenarios fall clearly into one

of the categories; however, naturally, some cross the boundaries between categories. The types of use cases identified are summarised below.

3.1 Data Annotation

The Data Annotation use cases comprise a broad range of scenarios involving human annotation of the emotion contained in some data, e.g. speech samples or video clips. These scenarios show a broad range with respect to the material being annotated, the way this material is collected, the way the emotion itself is represented, and, notably, which kinds of additional information about the emotion are being annotated.

One simple case is the annotation of plain text with emotion dimensions, notably valence, as well as with emotion categories and intensities. Recent work on naturalistic multimodal emotional recordings has compiled a much richer set of annotation elements (Douglas et al. 2090), and has argued that a proper representation of these aspects is required for an adequate description of the inherent complexity in naturally occurring emotional behaviour. Examples of such additional annotations are multiple emotions that co-occur in various ways (e.g., as blended emotions, as a quick sequence, as one emotion masking another one), regulation effects such as simulation or attenuation, confidence of annotation accuracy, or the description of the annotation of one individual versus a collective annotation. Data is often recorded by actors rather then observed in naturalistic settings. Here, it may be desirable to represent the quality of the acting, in addition to the intended and possibly the perceived emotion. With respect to requirements, Data Annotation poses the most complex kinds of requirements towards an emotion markup language, because many of the subtleties humans can perceive are far beyond the capabilities of today's technology.

3.2 Emotion Recognition

The general context of the Emotion Recognition use case has to do with low- and mid-level features which can be automatically detected, either offline or online, from human-human and human-machine interaction. In the case of low-level features, these can be facial features, such as Action Units (AUs) (Ekman and Friesen 1978) or MPEG 4 facial action parameters (FAPs) (Tekalp and Ostermann 2000), speech features related to prosody (Devillers et al. 2005) or language, or other, less frequently investigated modalities, such as bio signals, like heart rate or skin conductivity (e.g. (Picard et al. 2001)). All of the above can be used in the context of emotion recognition to provide emotion labels or extract emotion-related cues, such as smiling, shrugging or nodding, eye gaze and head pose, etc. These features can then be transferred into higher levels of abstraction, stored for further processing

or reused to synthesise expressivity in an embodied conversational agent (ECA) (Bevacqua et al. 2006).

In the case of unimodal recognition, the most prominent examples are speech and facial expressivity analysis. Regarding speech prosody and language, the CEICES data collection and processing initiative (Batliner et al. 2006) as well as exploratory extensions to automated call centres (Burkhardt et al. 2005) are the main factors that define the essential features and functionality of this use case.

Furthermore, individual modalities can be merged, either at feature- or decision-level, to provide multimodal recognition (Castellano et al. 2008) (McIntyre and Göcke 2008). A fusion on the feature level means that the classifier combines features from several modalities and decides on the basis of the unified features. An alternative would the classification (or recognition) based on the outcome of several classifiers, one for each modality (decision level). Features and timing information (duration, peak, slope, etc.) from individual modalities are still present, but an integrated emotion label is also assigned to the multimedia file or stream in question. In addition to this, a confidence measure for each feature and decision assists in providing flexibility and robustness in automatic or user-assisted methods.

3.3 Emotion Generation

The use cases in the Generation category were divided into three further sub categories, dealing with the simulation of modelled emotional processes, the generation of face and body gestures and the generation of emotional speech.

The first sub set of generation use cases are termed 'Affective Reasoner', to denote emotion modelling and simulation. The use cases in this category have a number of common elements that represented triggering the generation of an emotional behaviour according to a specified model or mapping. In general, emotion eliciting events are passed to an emotion generation system that maps the event to an emotion state which could then be realised as a physical representation, e.g. as gestures, speech or behavioural actions as described in the use cases of the other two sub sets.

The second sub set deals with the generation of automatic facial and body gestures for characters. With these use cases, the issue of the range of possible outputs from emotion generation systems becomes apparent. While all focused on generating human facial and body gestures, the possible range of systems that they connect to is large, meaning the possible mappings or output schema would be large. Both software and robotic systems are represented and as such the generated gesture information could be sent to both software and hardware based systems on any number of platforms. While a number of standards are available for animation that are used extensively within academia (e.g., MPEG-4 (Tekalp and Ostermann 2000), BML (Kopp et al. 2006)), they are by no means common in industry.

The final sub set is primarily focused on issues surrounding emotional speech synthesis, dialogue events and paralinguistic events. Similar to the previous sub sets, this is also complicated by the wide range of possible systems to which the gener-

ating system will pass its information. There does not seem to be a widely used common standard, even though the range is not quite as diverse as with facial and body gestures. Some of these systems make use of databases of emotional responses and as such might use an emotion language as a method of storing and retrieving this information.

4 Requirements for an emotion markup language

How useful is a markup language for emotions? This crucially depends on its ability to provide the representations required in a given use case.

The Emotion Incubator Group has analysed the above-mentioned use cases in order to make the implicit requirements contained in the descriptions explicit; to structure them in a way that reduces complexity; and to agree on the boundaries between what should be included in the language itself, and where suitable links to other kinds of representations should be used. Given the thematic breadth of use cases, the integration of requirements from all the use cases into one coherent document was not straightforward. For example, many application domains have their own specific kinds of expressive behaviour that are important and that must be linked with the emotion representation. Also, what is called "input" in the case of recognising emotions, such as a facial expression, would be called "output" in the case of generation, and vice versa. Therefore, to delimit the area to be covered by the emotion markup language, and in order to find a common vocabulary that can be used across application domains, two basic principles were agreed.

1. The emotion language should not try to represent sensor data, facial expressions, etc., but define a way of interfacing with external representations of such data.
2. The use of system-centric vocabulary such as "input" and "output" should be avoided. Instead, concept names should be chosen by following the phenomena observed, such as "experiencer", "trigger", or "observable behaviour".

The resulting requirements are reported in detail in the Final Report of the W3C Emotion Incubator Group (Schröder 2007); the key elements are summarised in the following.

4.1 Core emotion description

For the emotion (or emotion-related state) itself, three types of representation are envisaged, which can be used individually or in combination.

- Emotion categories (words) are symbolic shortcuts for complex, integrated states; an application using them needs to take care to define their meaning properly in

the application context. The emotion markup should provide a generic mechanism to represent broad and small sets of possible emotion-related states. It should be possible to choose a set of emotion categories (a label set), because different applications need different sets of emotion labels.

- Alternatively, or in addition, emotion can be represented using a set of continuous dimensional scales, representing core elements of subjective feeling and of people's conceptualisation of emotions. As for emotion categories, it is not possible to predefine a normative set of dimensions. Instead, the language should provide a "default" set of dimensions, such as arousal, valence and power, that can be used if there are no specific application constraints, but allow the user to "plug in" a custom set of dimensions if needed.
- As a third way to characterise emotions and related states, appraisal scales can be used, which provide details of the individual's evaluation of their environment. Examples include novelty, goal significance, or compatibility with one's standards.
- Finally, it may potentially be relevant to characterise emotions in terms of the action tendencies that accompany them, such as approach or avoidance.

Other concepts that appear to be necessary for describing the "core" of an emotion include:

- the intensity of an emotion;
- the representation of multiple and complex emotions, in cases where emotions may be co-occurring (such as being sad about one thing and angry about another at the same time), or in cases of regulation;
- regulation, i.e. the complex of phenomena where an experiencer modifies the emotion itself or its expression, e.g. by masking, inhibiting, simulating, etc.;
- and temporal aspects of the emotion. A generic mechanism for temporal scope should allow different ways to specify temporal aspects, such as start and end times, start time and duration, timing relative to another entity (e.g., "start 2 seconds before an utterance starts and end with the second noun-phrase..."). In addition, a sampling mechanism providing values for a parameter at evenly spaced time intervals would allow for the description of continuous values such as intensity or dimensional scales as they change over time[1].

4.2 Meta information about emotion description

Three additional requirements with respect to meta information have been elaborated:

[1] Note that the timing of any associated behaviour, triggers etc. is covered in section "Links to the rest of the world"...

1. information concerning the degree of acting of emotional displays; so, a mechanism should be defined to add special attributes for acted emotions such as perceived naturalness, authenticity, quality, etc.;
2. information related to confidences and probabilities of emotional annotations. The emotion markup should provide a generic attribute for representing the confidence (or, inversely, uncertainty) of any aspect of the representation. For example, such an attribute can be used to reflect the confidence of a human annotator that the particular value is as stated (e.g., a confidence of 0.8 that a subject's expression corresponds to the emotion "happiness"). More generally, it may be necessary to represent such a confidence with respect to each level of representation: intensity, degree of acting, etc.;
3. and finally the modalities involved (e.g. face, voice, body posture or hand gestures, but also lighting, font shape, etc.).

All of this information thereby applies to each annotated emotion separately.

4.3 Links to the "rest of the world"

In order to be properly connected to the kinds of data relevant in a given application scenario, several kinds of links are required, referring to external media objects or to a position on a time-line within a media file. Start and end times are important to mark onset and offset of an emotional episode. Relevant information to link to can be of various sorts, so that a mechanism should be defined for flexibly assigning meaning to a link. A reasonable initial set of meanings for such links to the "rest of the world" should include: the experiencer, i.e. the person who "has" the emotion; the observable behaviour "expressing" it; the trigger, cause, or eliciting event of an emotion; and the object or target of the emotion, that is, what the emotion is about.

4.4 Global metadata

Representing emotion, be it for annotation, detection or generation, requires the description of the context not directly related to the description of emotion per se but also the description of a more global context which is required for exploiting the representation of the emotion in a given application. Examples are data on person(s) like ID, date of birth, gender, language, personality traits, culture or level of expertise as labeller, information about the intended application (e.g. purpose of classification; application type - call centre data, online game, etc.; possibly, application name and version) and furthermore, it should be possible to specify the technical environment, for example, links to the particular camera properties, sensors used (model, configuration, specifics), or indeed any kind of environmental data. Finally, information on the social and communicative environment will be required, such as

the type of collected data or the situational context in which an interaction occurs (number of people, relations, link to description of individual participants).

5 Different descriptive schemes for emotions

5.1 Glancing at theories

The collection of use cases and subsequent definition of requirements presented so far was performed in a predominantly bottom-up fashion, and thus provides a strongly application centered, engineering driven perspective. The purpose of this section is to put this into a more theory centered perspective. A representation language should be as theory independent as possible but by no means ignorant of psychological theories. Therefore a crosscheck to which extent components of existing psychological models of emotion are mirrored in the currently collected requirements is performed.

The old Indian tale of the blind men and the elephant gained some popularity in the psychological literature as an allegory for the conceptual difficulties to come up with unified and uncontroversial descriptions of complex phenomena such as emotions. In this tale several blind men, who never have encountered an elephant before, try to come up with an understanding of the nature of this unknown object. Depending on the body part each of them touches they provide strongly diverging descriptions. An elephant seems to be best described as a rope if you hang to its tail only, is an ensemble of columns if you just touched its legs, appears as a piece of cloth if you encountered its ears etc.

This metaphor fits nicely with the multitude of definitions and models currently available in the scientific literature on emotions, which come with a fair amount of terminological confusion added on top. Cowie et al. (Cowie et al. 2090) give a very good overview. There are no commonly accepted answers to the questions on how to model the underlying mechanism that are causing emotions, on how to classify them, on whether to use categorical or dimensional descriptions etc. But leaving these questions aside, there is a core set of concepts that are quite readily accepted to be essential components of emergent emotions.

One terminological issue quite relevant for the discussion in this section is the semantics of the term emotion itself, which has been used in a broad and a narrow sense.

In its narrow sense, used e.g. by Scherer, the term refers to what is also called a prototypical emotional episode (Russell and Feldman 1999), full blown emotion, emergent emotion (Cowie et al. 2090): a short, intensive, clearly event triggered emotional event, of which fear when encountering a bear in the woods and fleeing in terror is the favourite example.

Especially in technological contexts there is a tendency to use the term emotion(al) in a broad sense, in its extreme for almost everything that cannot be cap-

tured as a purely cognitive aspect of human behaviour. More established but still not concisely defined terms for the range of phenomena that make up the elements of emotional life are “emotion-related states”, “affective states” and “pervasive emotions”. Whatever term used, there is quite some agreement that apart from emergent emotions the group of affective states includes moods, interpersonal stances, preferences/attitudes, and affect disposition.

The envisaged scope of an emotion representation language clearly is concerned with emotions in the broad sense, i.e. it should be able to deal with different emotion-related states. Emergent emotions - not without reason also termed prototypical emotional episodes - can be viewed as the archetypical affective states and many emotional theories focus on emergent emotions. Empirical studies (Cowie et al. 2090) (Wilhelm et al. 2004) on the other hand indicate that while there are almost no instances where people report their state as completely unemotional, examples of full-blown emergent emotions are really quite rare. As the majority of the ever present emotional life consists of moods, stances towards objects and persons, and altered states of arousal, these naturally should play an increasingly prominent role in emotion-related computational applications and are thus clearly in the scope of the representation language.

5.2 Core concepts and their role in the representation language

As stated above, although there is much disagreement on how best to theoretically model emotions, there is a certain consensus on a number of components that do play a role in the emotional life. The following list presents prominent concepts that have been used by psychologists in their quest for describing emotions. It will be evaluated whether and how these concepts are mirrored in the current list of requirements.

A general heuristic in the design of a representation language for emotions is to focus on those concepts that are observable in some way, thus hidden processes and constructs that are defined conceptually and not experientially should not be part of the representation.

Subjective component: Feelings. Feelings have not been mentioned in the requirements at all. They are not to be explicitly included in the representation for the moment being, as they are defined as internal states of the subject and are thus not accessible to observation. Applications can be envisaged where feelings might be of relevance in the future though, e.g. if self-reports are to be encoded. It should thus be kept as an open issue whether to allow for an explicit representation of feelings as a separate component in the future.

Cognitive component: Appraisals. As a reference to appraisal-related theories, the OCC model (Ortony et al. 1988) which is especially popular in the computational domain, has been brought up in the use cases. In these models emotions are elicited by a cognitive evaluation of perceived events or situations by a number of checks along different dimensions (e.g. relevance, coping potential). No choice for

the exact set of appraisal conditions is to be made here. An open issue is whether models that make explicit predictions on the temporal ordering of appraisal checks (Sander et al. 2005) should be encodable to that level of detail. In general, appraisals are to be encoded in the representation language via attributing links to trigger objects. The encoding of other cognitive aspects, i.e. effects of emotions on the cognitive system (memory, perception, etc.) is to be kept as an open issue.

Physiological component: Changes in heart-rate, breathing, sweating etc. obviously are a component of emergent emotions. They also are quite interconnected with other components in this list. Feelings, e.g., can just be conceived as perception and classification of these physiological states, as in (James 1984). Physiology, e.g. changes in the muscular tone, also account for changes of expressive features in speech (prosody, articulatory precision) or in the appearance (posture, skin colour). Physiological measures have been mentioned in the context of the use case of emotion recognition. They are to be integrated in the representation via links to externally encoded measures conceptualised as "observable behaviour".

Behavioural component: Action tendencies, such as the tendency to approach, avoid, or reject are a central concept in the work of Frijda (Frijda 1986). Action tendencies can be viewed as a link between the outcome of an appraisal process and actual actions. It remains an issue of theoretical debate whether action tendencies, in contrast to actions, are among the set of actually observable concepts. Nevertheless these should be integrated in the representation language. This once again can be achieved via the link mechanism, this time an attributed link can specify an action tendency together with its object or target.

Expressive component: Expressions, which are most frequently studied in the face, but also as signs in the voice, posture and gesture, play a central role in the research tradition that relies on so called basic emotions, e.g. in the work of Ekman (Ekman 1992). But for the sake of the representation language no stance needs to be taken whether expressions are produced by innate mechanisms as proposed by evolutionary models or are fundamentally communicative. Expressions are frequently referred to in the requirements. There is agreement to not encode them directly but again to make use of the linking mechanisms to observable behaviours.

Dimensional descriptions somehow cut across different components mentioned so far. There is a rich tradition of models that describe emotions in terms of a combination of typically two or three dimensions going back as far as Wundt (Wundt 1905). Though there are differences in the concrete terminology, valence (positive or negative, pleasantness) and activation (arousal, energy) are rather undisputed dimensions. Dimensions, however, are not only used as global descriptors for emotions; typically, appraisals are formulated in terms of dimensions as well. Russell and Feldman Barrett (Russell and Feldman 1999) use dimensions for describing feelings (core-affects in their terminology), and there are studies where dimensions are used for statistically clustering lexical terms for emotions in different languages (Mehrabian 1995) (Roesch et al. 2006). For the design of a representation language it can thus be concluded that expressive means for dimensional encodings should be made quite generally available.

Category based descriptions, i.e. systems that make use of single terms such as anger or fear for representing emotions, are appealingly simple and are of course a prominent option in the envisaged representation. They are not without problems though. As Cowie et al. (Cowie et al. 2090) spell out, these terms often lack concise definitions, might be too unspecific, obscure the influence of the context etc. In a theoretical context categorical descriptions are typically connected to theories that tie emotions to basic innate neural systems and claim that there is a small number of such basic emotions that are nicely reflected in everyday natural language terms. In the context of a representation language for computational purposes, this strong commitment to a certain psychological theory plays a less important role. Even when using simple terms like fear, it still can be conceived as just a handy abbreviation for a complex process, where a subject encountering a bear in the woods initiates the evaluation of a cascade of appraisal dimensions, experiences strong changes in their physiological states, their perceptive and cognitive capabilities, feelings of fear and develops not only an action tendency for fleeing but also performs this action, while crying out in panic. Also, another researcher might use the label fear for quite a different experience, for instance a user realising loss of sensible data. Good use of meta information are hence required when using category based descriptions, accurately describing the situation in which the emotion occurred.

Figure 1 summarises the way in which these components of emotion are related to the requirements for an emotion markup language as summarised in Section 6.4.4.

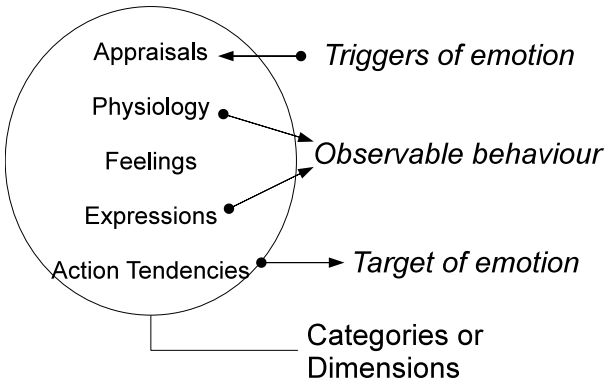

Fig. 1 Overview of how components of emotions are to be linked to external representations.

5.3 Emergent emotions vs. other emotion-related states

In the start of this section it was elaborated that the scope of the representation language should not be restricted to emergent emotions even though these have received most attention so far. A positive aspect is that, although emergent emotions

make up only a very small part of the emotion-related states, they nevertheless are sort of archetypes. Representations developed for emergent emotions should be usable as a basis for the encoding of other important emotion-related states such as moods and attitudes.

Scherer (Scherer 2000), who systematically defines relationships between emergent emotions and other emotion-related states, supports this hypothesis. Scherer is proposing a small set of so-called design features for characterising the various kinds of affective state: intensity, duration, synchronisation, event focus, appraisal elicitation, rapidity of change, and behavioural impact. For example, emergent emotions are defined as having a strong direct impact on behaviour, high intensity, being rapidly changing and short, are focussing on a triggering event and involve strong appraisal elicitation. Moods, as another kind of emotion-related states, are described using the same set of design features: they are characterised as not having a direct impact on behaviour, being less intense, changing less quickly and lasting longer, and not being directly tied to an eliciting event. In this framework, different types of emotion-related states thus just arise from differences in the design features. In technological contexts, however, descriptions will rarely reach this level of detail. Therefore, the representation language should provide simple means for encoding the type of an emotion related state explicitly.

6 Options for realisation in XML

This section addresses the question how a generic language to annotate and represent emotions should be realised syntactically. Such a language should be (i) easily combined with other markup languages, and (ii) extensible in order to adapt it to specific domains.

Several options are available. Given the fact that most recent markup languages are defined in XML (eXtensible Markup Language), it appears to be a good choice to formulate an emotion markup language in XML. An alternative to this is the use of RDF (Resource Description Framework), a formalism for representing information as subject-predicate-object expressions. RDF is particularly well-suited for representing ontological structures, i.e. relations among entities. An XML formulation of RDF exists, making the two alternatives non-exclusive. Another alternative is the use of OWL (Ontology Web Language) which facilitates the representation of ontologies (cf. Section 6.4.7) by providing additional vocabulary along with a formal semantics. OWL adds more vocabulary for describing properties, classes of concepts and their interrelationships. It is based on XML and RDF. Without aiming to prejudge this question, the following discussion will focus on XML realisation.

6.1 Flat vs. deep structures

The structure of annotation in a markup language can be flat or deep. Deep structures have the advantage that the meaning of information is fully explicit in the structure; however, this adds overhead which becomes a burden especially in simple cases. Flat structures, on the other hand, can represent simple cases in a simple way, but may become difficult to read for more complex annotations.

The HUMAINE EARL (Emotion Annotation and Representation Language) [24] is an example of a flat structure. Its design was guided by the principle that "simple cases should look simple". Indeed, the annotation of a picture with an emotion category and intensity would be as simple as the following:

Example 5. Emotion category and intensity in EARL

<emotion xlink:href="picture.jpg" category="contentment" intensity="0.7" />

When this annotation is to be enriched with additional annotations, such as the appraisal "high goal conduciveness" and the dimensional ratings "positive and passive", these are simply added to the list of attributes:

Example 6. Additional annotations in EARL

<emotion xlink:href="picture.jpg" category="contentment" intensity="0.7"
goal_conduciveness="0.9" valence="0.5" arousal="-0.3" />

In such a structure, the meaning of the various attributes is implicit - the user needs to know what they being doing, or else it is easy to mix things in an invalid way.

A deep structure could help avoid this problem, e.g. by making the status of category, dimension or appraisal of an annotation explicit, e.g.:

Example 7. Deep structure in EARL

<emotion>
<object xlink:href="picture.jpg" />
<category intensity="0.7">contentment</category>
<appraisal>
<goal_conduciveness>0.9</goal_conduciveness>
</appraisal>
<dimensions>
<valence>0.5</valence>
<arousal>-0.3</arousal>
</dimensions>
</emotion>

In this example, the trade-off between simplicity and explicit structures becomes clearly visible. The disambiguation of annotations is paid for by a more complicated structure.

Which alternative is better to use remains to be seen. As both alternatives have all their information embedded in one ¡emotion¿ tag, there is no difference regarding their capabilities for being combined with other markup languages.

7 Extensibility

The question of extensibility is a challenge with respect to the definition of any standard. In the present case, where very different emotion representations are suitable in different circumstances, and sets of values are often domain-specific, the challenge is particularly marked.

The simplest solution to providing this kind of flexibility would be to leave the sets of values open. It would provide maximum flexibility, but on the other hand would make it impossible to verify that the markup is correct and meaningful.

The HUMAINE EARL specification has proposed a more controllable method for defining custom sets of values. It allows users to "plug in" their own, tailor-made sets of emotion categories, dimensions and/or appraisals. Technically, this is achieved by splitting the EARL schema design in four parts: for any given EARL "dialect", the EARL base schema, which defines the structure of EARL documents, is complemented by three small schemas defining the set of categories, dimensions and/or appraisals to be used. In this way, the EARL is actually conceived as a family of EARL dialects - all of them sharing a common structure, but each with its own set of valid values and identified by its own XML namespace. As a "default" EARL language, a set of 48 categories, three dimensions and 19 appraisals were proposed. These can be used when there are no specific requirements to go beyond them.

An alternative approach may be for the markup to indicate, as an attribute value, a namespace to be used for validating substructures. This approach was used in the W3C working draft language EMMA (extensible multimodal annotation) in its ¡emma:model¿ mechanism: A namespace reference can be used to indicate the possible substructures of an ¡emma:interpretation¿ element. This mechanism appears to be more flexible, but also to introduce more overhead compared to the HUMAINE EARL approach.

8 Ontologies of emotion descriptions

In an emotion representation language, different emotion representations need to be made possible because no preferred representation has yet emerged for all types of use. Instead, the most suitable representation to use depends on the application. As a result, complex systems such as many foreseeable real-world applications will require some information about (1) the relationships between the concepts used in one description, and about (2) the relationships between different descriptions. In order to enable components in complex systems to work together even though they use different emotion representations, an emotion markup language should be complemented with a mapping mechanism based on ontologies of emotion descriptions.

The concepts in an emotion description are usually not independent, but are related to one another. For example, emotion words may form a hierarchy, as suggested e.g. by prototype theories of emotions. For example, Shaver et al. (Shaver et al. 1987) classified cheerfulness, zest, contentment, pride, optimism enthrallment and relief

as different kinds of joy, irritation, exasperation, rage, disgust, envy and torment as different kinds of anger, etc.

Such structures, be they motivated by emotion theory or by application-specific requirements, may be an important complement to the representations in an emotion markup language. In particular, they would allow for a mapping from a larger set of categories to a smaller set of higher-level categories.

Different emotion representations (e.g., categories, dimensions, and appraisals) are not independent; rather, they describe different parts of the "elephant", of the phenomenon emotion. However, from a scientific point of view, it will not always be possible to define such mappings. For example, the mapping between categories and dimensions will only work in one direction. Emotion categories, understood as short labels for complex states, can be located on emotion dimensions representing core properties; but a position in emotion dimension space is ambiguous with respect to many of the specific properties of emotion categories, and can thus only be mapped to generic super-categories.

Similarly, it may be possible to define a mapping from categories to appraisals, when categories are understood as "shortcuts" for appraisal configurations. In the opposite direction, however, many appraisal combinations will not be associated with any category in an exact manner; instead, they may be "similar" to one or more categories.

The mapping mechanism required here could easily become as complex as the field of emotion theory itself, and the attempt to define such mappings could end up as an interminable discussion about theoretical notions. Pragmatically, however, a subset of possible mappings may be defined, at least in application-specific ways, and the concrete needs of applications may be a suitable guideline for the definition of a mapping mechanism, helping to avoid the pitfall of getting stuck in theoretical debate.

9 Conclusion and Outlook

This chapter has described a number of basic considerations that should be addressed by an emotion markup language. We have briefly reviewed existing work, pointing out the potential benefits of a reusable standard "plug in" representation of emotions and related states. We have described the compilation of a rich set of requirements from use cases in the W3C Emotion Incubator Group, and have compared the work with scientific descriptions of emotion. This comparison has shown that many aspects studied in the literature are also relevant for the technological use cases envisaged, whereas some aspects of high scientific importance, such as feeling, play only a limited role. Finally, we have pointed out basic design choices available for the syntactic realisation of an emotion representation, notably in terms of a flat or deep structure, and have pointed out the potential use of emotion ontologies for the interoperability of components using different types of emotion descriptors.

The considerations described here are a solid basis for the development of an emotion markup specification. While various aspects may need to be simplified in view of implementability, the present collection is valuable as an outline of the actual complexity of the phenomenon and should be able to serve as a guideline for future standardisation activities in the area of representing emotions and related states.

References

[Batliner et al. 2006] Batliner A, Steidl S, Schuller B, Seppi D, Laskowski K, Vogt T, Devillers L, Vidrascu L, Amir N, Kessous L, and Aharonson V (2006). Combining efforts for improving automatic classification of emotional user states. In Erjavec T and Gros J (eds), Language Technologies, IS-LTC 2006, pp 240-245. Ljubljana, Slovenia: Infornacijska Druzba (Information Society).

[Bevacqua et al. 2006] Bevacqua E, Raouzaiou A, Peters C, Caridakis G, Karpouzis K, Pelachaud C and Mancini M (2006) Multimodal sensing, interpretation and copying of movements by a virtual agent. In André E, Dybkjaer L, Minker W, Neumann H, and Weber M (eds): Proceedings of Perception and Interactive Technologies (PIT'06), pp 164-174, Kloster Irsee, Germany. Springer, LNCS 4021. doi: 10.1007/11768029_16.

[Burkhardt et al. 2005] Burkhardt F, van Ballegooy M, Englert R and Huber R (2005) An Emotion-Aware Voice Portal, In: Proc. Electronic Speech Signal Processing ESSP 2005, Prague, pp 123131.

[Castellano et al. 2008] Castellano G, Kessous L and Caridakis G (2008) Emotion recognition through multiple modalities: face, body gesture, speech. In: Peter C and Beale R (eds): Affect and Emotion in Human-Computer Interaction. LNCS, vol 4868. Springer, Heidelberg.

[Cowie et al. 2090] Cowie R, Sussman N, and Ben-Ze'ev A (this book) Emotions: concepts and definitions.

[de Carolis et al. 2004] de Carolis B, Pelachaud C, Poggi I and Steedman M (2004) APML, a Mark-up Language for Believable Behavior Generation, In Helmut Prendinger (ed): Life-like Characters. Tools, Affective Functions and Applications, Springer.

[Devillers et al. 2005] Devillers L, Vidrascu L, and Lamel L (2005) Challenges in real-life emotion annotation and machine learning based detection. Neural Networks 18:407-422.

[Douglas et al. 2090] Douglas-Cowie E, Cox C, Lowry O, Martin J-C, Devillers L, Abrilian S, Pelachaud C, Peters C (this book) The HUMAINE Database.

[Ekman and Friesen 1978] Ekman P and Friesen W (1978) The Facial Action Coding System. Consulting Psychologists Press, San Francisco.

[Ekman 1992] Ekman P (1992) An argument for basic emotions. Cognition and Emotion 6:169-200.

[Frijda 1986] Frijda N (1986) The Emotions. Cambridge: Cambridge University Press.

[James 1984] James W (1884) What is an Emotion? Mind 9:188-205.

[Gustavsson et al. 2001] Gustavsson C, Beard S, Strindlund L, Huynh Q, Wiknertz E, Marriott A, and Stallo J (2001) VHML Specification Working Draft v0.3, October 21st 2001. Retrieved 24 October 2007 from: `http://www.vhml.org/downloads/VHML/vhml.pdf`

[Kopp et al. 2006] Kopp S, Krenn B, Marsella S, Marshall A, Pelachaud C, Pirker H, Thórisson K and Vilhjálmsson H (2006) Towards a common framework for multimodal generation in ECAs: The Behavior Markup Language. In: Proceedings of the 6th International Conference on Intelligent Virtual Agents (IVA'06), Marina del Rey, USA, pp 205-217.

[Krenn et al. 2002] Krenn B, Pirker H, Grice M, Piwek P, Deemter K van, Schröder M, Klesen M, and Gstrein E (2002) Generation of multimodal dialogue for net environments. In Busemann S (ed): Proceedings of Konvens. Saarbrücken, Germany.

[McIntyre and Göcke 2008] McIntyre G and Göcke R (2008) The Composite Sensing of Affect. In: Peter C and Beale R (eds): Affect and Emotion in Human-Computer Interaction. LNCS, vol. 4868. Springer, Heidelberg

[Mehrabian 1995] Mehrabian A (1995) Framework for a comprehensive description and measurement of emotional states. Genetic, Social, and General Psychology Monographs 121:339-361.

[Ortony et al. 1988] Ortony A, Clore GL and Collins A (1988) The cognitive structure of emotions. New York: Cambridge University Press.

[Picard et al. 2001] Picard RW, Vyzas E and Healey J (2001) Toward Machine Emotional Intelligence - Analysis of Affective Physiological State. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(10):1175-1191

[Roesch et al. 2006] Roesch EB, Fontaine JB and Scherer KR (2006) The world of emotion is two-dimensional - or is it? Paper presented to the HUMAINE Summer School 2006, Genoa, Italy.

[Russell and Feldman 1999] Russell JA and Feldman BL (1999) Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant. Journal of Personality and Social Psychology 76:805-819.

[Sander et al. 2005] Sander D, Grandjean D, and Scherer KR (2005) A systems approach to appraisal mechanisms in emotion. Neural Networks 18:317-352.

[Scherer 2000] Scherer KR (2000) Psychological models of emotion. In Borod JC (ed): The Neuropsychology of Emotion, pp 137-162. New York: Oxford University Press.

[Schröder 2006] Schröder M, Pirker H and Lamolle M (2006) First suggestions for an emotion annotation and representation language. In: Proceedings of LREC'06 Workshop on Corpora for Research on Emotion and Affect, Genoa, Italy, pp 88-92

[Schröder 2007] Schröder M, Zovato E, Pirker H, Peter C, and Burkhardt F (2007). W3C Emotion Incubator Group Final Report, Published online on

10 July 2007: http://www.w3.org/2005/Incubator/emotion/XGR-emotion/.

[Shaver et al. 1987] Shaver P, Schwartz J, Kirson D, and O'Connor C (1987). Emotion knowledge: Further exploration of a prototype approach. Journal of Personality and Social Psychology 52:1061-1086.

[Tekalp and Ostermann 2000] Tekalp M and Ostermann J (2000) Face and 2d mesh animation in MPEG-4. Image Communication Journal 15:387-421.

[Wilhelm et al. 2004] Wilhelm P, Schoebi D and Perrez M (2004) Frequency estimates of emotions in everyday life from a diary method's perspective: a comment on Scherer et al.'s survey-study "Emotions in everyday life". Social Science Information 43(4):647-665.

[Wundt 1905] Wundt W (1905) Grundriss der Psychologie [Fundamentals of psychology]: Vol 3 (5th ed.), Leipzig: Engelmann.