What should a generic emotion markup language be able to represent?

Marc Schröder[1], Laurence Devillers[2], Kostas Karpouzis[3], Jean-Claude Martin[2], Catherine Pelachaud[4], Christian Peter[5], Hannes Pirker[6], Björn Schuller[7], Jianhua Tao[8], and Ian Wilson[9]

[1] DFKI GmbH, Saarbrücken, Germany
[2] LIMSI-CNRS, Paris, France
[3] Image, Video and Multimedia Systems Lab, Nat. Tech. Univ. Athens, Greece
[4] Univ. Paris VIII, France
[5] Fraunhofer IGD, Rostock, Germany
[6] OFAI, Vienna, Austria
[7] Tech. Univ. Munich, Germany
[8] Chinese Acad. of Sciences, Beijing, China
[9] Emotion AI, Tokyo, Japan
http://www.w3.org/2005/Incubator/emotion

Abstract. Working with emotion-related states in technological contexts requires a standard representation format. Based on that premise, the W3C Emotion Incubator group was created to lay the foundations for such a standard. The paper reports on two results of the group's work: a collection of use cases, and the resulting requirements. We compiled a rich collection of use cases, and grouped them into three types: data annotation, emotion recognition, and generation of emotion-related behaviour. Out of these, a structured set of requirements was distilled. It comprises the representation of the emotion-related state itself, some meta-information about that representation, various kinds of links to the "rest of the world", and several kinds of global metadata. We summarise the work, and provide pointers to the working documents containing full details.

1 Introduction

As emotion-oriented computing systems are becoming a reality, the need for a standardised way of representing emotions and related states is becoming clear. For real-world human-machine interaction systems, which typically consist of multiple components covering various aspects of data interpretation, reasoning, and behaviour generation, it is evident that emotion-related information needs to be represented at the interfaces between system components.

The present paper reports on a joint effort to lay the basis for a future standard for representing emotion-related states in a broad range of technological contexts. After briefly revisiting previous work, we introduce the W3C Emotion Incubator group, before we describe two of its key results: a rich collection of use

cases – scenarios where an emotion markup language would be needed –, and a compilation of the requirements resulting from these use cases.

1.1 Previous work

Until recently, when markup languages provided for the representation of emotion, it was part of a more complex scenario such as the description of behaviour for embodied conversational agents (ECAs) [1]. The expressivity of the representation format was usually very limited – often, only a small set of emotion categories was proposed, such as the "big six" which according to Ekman [2] have universal facial expressions, and their intensity. When additional descriptions of an emotion were offered, these were closely linked to the particular context in which the language was to be used. As a result, these languages cannot generally be used outside the specific application for which they were built.

Two recent endeavours have proposed more comprehensive descriptions of emotion-related phenomena. The Emotion Annotation and Representation Language (EARL – [3]), developed in the HUMAINE network on emotion-oriented computing, has made an attempt to broaden the perspective on representing emotion-related information. The EARL is a syntactically simple XML language designed specifically for the task of representing emotions and related information in technological contexts. It can represent emotions as categories, dimensions, or sets of appraisal scales. As different theories postulate different sets of emotion words, dimensions and appraisals, the design is modular, so that the appropriate set of descriptors for the target use can be chosen. In addition, a set of attributes can represent intensity and regulation-related information such as the suppression or simulation of emotion. Complex emotions, which consist of more than one "simple" emotion, can also be represented. A detailed specification including an XML schema can be found at `http://emotion-research.net/earl`.

The HUMAINE database annotation scheme, developed independently of the EARL, has a slightly different focus. The HUMAINE team working on databases explored the annotation of a variety of emotional samples collected from different types of databases including induced, acted and naturalistic behaviours. A modular coding scheme [4] was defined to cover the requirements coming from these different data. This scheme enables the description at multiple levels of the emotional content and was applied to the annotation of French and English TV interviews. It is defined as a structured set of modular resources from which researchers can select what they need to match their own research requirements for the annotation of emotional data:

- Global emotion descriptors, used for representing emotion perceived in a whole clip: emotion words, emotion related states (e.g. attitudes), combination types, authenticity, core affect dimensions, context labels, key events and appraisal categories;
- Emotion descriptors varying over time: eight dimensional traces, such as the perceived variation of the level of acting during the clip;
- Signs of emotion: speech and language, gesture and face descriptors.

The conceptual coding scheme is implemented in XML in the Anvil tool format and is available for download from the HUMAINE web site[1].

1.2 The W3C Emotion Incubator group

The W3C Emotion Incubator group (`http://www.w3.org/2005/Incubator/emotion`) was created to investigate the prospects of defining a general-purpose Emotion annotation and representation language. The group consists of representatives of 15 institutions from 11 countries in Europe, Asia, and the US. The approach chosen for the group's work has been to revisit carefully the question where such a language would be used, and what those use case scenarios require from a language, before even starting to discuss the question of a suitable syntactic form for the language. In the following, the result of these two working steps are summarised.

2 Use cases

With the Emotion Incubator group taking a solid software engineering approach to the question of how to represent emotion in a markup language the first necessary step was to gather together as complete a set of use cases as possible for the language. At this stage, we had 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. The resulting set of final use cases would then be used as the basis for the next stage of the design processes, the definition of the requirements of the language.

The Emotion Incubator group is comprised of people with wide ranging interests and expertise in the application of emotion in technology and research. Using this as a strength, we asked each member to propose one or more use case scenarios that would represent the work they, themselves, were doing. This allowed the group members to create very specific use cases based on their own domain knowledge. Three broad categories were defined for these use cases: Data Annotation, Emotion Recognition and Emotion Generation. Where possible we attempted to keep use cases within these categories, however, naturally, some crossed the boundaries between categories.

A wiki was created to facilitate easy collaboration and integration of each member's use cases[2]. In this document, subheadings of the three broad categories were provided along with a sample initial use case that served as a template from which the other members entered their own use cases and followed in terms of content and layout. In total, 39 use cases were entered by the various working group members: 13 for Data Annotation, 11 for Emotion Recognition and 15 for Emotion Generation.

---

[1] `http://emotion-research.net/download/pilot-db`
[2] `http://www.w3.org/2005/Incubator/emotion/wiki/UseCases`

Possibly the key phase of gathering use cases was in the optimisation of the wiki document. Here, the members of the group worked collaboratively within the context of each broad category to find any redundancies (replicated or very similar content), to ensure that each use case followed the template and provided the necessary level of information, to disambiguate any ambiguous wording (including a glossary of terms for the project), to agree on a suitable category for use cases that might well fit into two or more and to order the use cases in the wiki so that they formed a coherent document.

In the following, we detail each broad use case category, outlining the range of use cases in each, and pointing out some of their particular intricacies.

2.1 Data annotation

The Data Annotation use case groups together a broad range of scenarios involving human annotation of the emotion contained in some material. 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. Similarly, simple emotional labels can be associated to nodes in an XML tree, representing e.g. dialogue acts, or to static pictures showing faces, or to speech recordings in their entirety. While the applications and their constraints are very different between these simple cases, the core task of emotion annotation is relatively straightforward: it consists of a way to define the scope of an emotion annotation and a description of the emotional state itself. Reasons for collecting data of this kind include the creation of training data for emotion recognition, as well as scientific research.

Recent work on naturalistic multimodal emotional recordings has compiled a much richer set of annotation elements [4], 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. In addition to annotations that represent fixed values for a certain time span, various aspects can also be represented as continuous "traces" – curves representing the evolution of, e.g., emotional intensity over time.

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, it has become clear that Data Annotation poses the most complex kinds of requirements with respect to an emotion markup language, because many of the subtleties humans can perceive are far beyond the

capabilities of today's technology. We have nevertheless attempted to encompass as many of the requirements arising from Data Annotation, not least in order to support the awareness of the technological community regarding the wealth of potentially relevant aspects in emotion annotation.

2.2 Emotion recognition

As a general rule, 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) [5] or MPEG 4 facial action parameters (FAPs) [6], speech features related to prosody [7] or language, or other, less frequently investigated modalities, such as biosignals (e.g. heart rate or skin conductivity). 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 stored for further processing or reused to synthesise expressivity on an embodied conversational agent (ECA) [8].

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 [9] as well as exploratory extensions to automated call centres are the main factors that defined the essential features and functionality of this use case. With respect to visual analysis, there are two cases: in the best case scenario, detailed facial features (eyes, eyebrows, mouth, etc.) information can be extracted and tracked in a video sequence, catering for high-level emotional assessment (e.g. emotion words). However, when analysing natural, unconstrained interaction, this is hardly ever the case since colour information may be hampered and head pose is usually not directed to the camera; in this framework, skin areas belonging to the head of the subject or the hands, if visible, are detected and tracked, providing general expressivity features, such as speed and power of movement [8].

For physiological data, despite being researched for a long time especially by psychologists, no systematic approach to store or annotate them is in place. However, there are first attempts to include them in databases [10], and suggestions on how they could be represented in digital systems have been made [11]. A main difficulty with physiological measurements is the variety of possibilities to obtain the data and of the consequential data enhancement steps. Since these factors can directly affect the result of the emotion interpretation, a generic emotion markup language needs to be able to deal with such low-level issues. The same applies to the "technical" parameters of other modalities, such as resolution and frame rate of cameras, the dynamic range or the type of sound field of the choosen microphone, and algorithms used to enhance the data.

Finally, individual modalities can be merged, either at feature- or decision-level, to provide multimodal recognition. In this case, 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.

2.3   Generation

We divided the 15 use cases in the generation category into a number of further sub categories, these dealt with essentially simulating modelled emotional processes, generating face and body gestures and generating emotional speech.

The use cases in this category had 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.

The generation use cases presented a number of interesting issues that focused the team on the scope of the work being undertaken. In particular, they showed how varied the information being passed to and information being received from an emotion processing system can be. This would necessitate either a very flexible method of receiving and sending data or to restrict the scope of the work in respect to what types of information can be handled.

The first sub set of generation use cases were termed 'Affective Reasoner', to denote emotion modelling and simulation. Three quite different systems were outlined in this sub category, one modelling cognitive emotional processes, one modelling the emotional effects of real time events such as stock price movements on a system with a defined personality and a large ECA system that made heavy use of XML to pass data between its various processes.

The next sub set dealt with the generation of automatic facial and body gestures for characters. With these use cases, the issue of the range of possible outputs from an emotion generation systems became apparent. While all focused on generating human facial and body gestures, the possible range of systems that they connect to was large, meaning the possible mappings or output schema would be large. Both software and robotic systems were 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 [6], BML [12]), they are by no means common in industry.

The final sub set was primarily focused on issues surrounding emotional speech synthesis, dialogue events and paralinguistic events. Similar to the issues above, the generation of speech synthesis, dialogue events, paralinguistic events etc. is complicated by the wide range of possible systems to which the generating 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 made use of databases of emotional responses and as such might use an emotion language as a method of storing and retrieving this information.

3 Requirements

Each use case scenario naturally contains a set of implicit "needs" or requirements – in order to support the given scenario, a representation format needs to be capable of certain things. The challenge with the 39 use case scenarios collected in the Emotion Incubator group was to make those implicit requirements explicit; to structure them in a way that reduces complexity; and to agree on the boundary between what should be included in the language itself, and where suitable links to other kinds of representations should be used.

Work proceeded in a bottom-up, iterative way. From relatively unstructured lists of requirements for the individual use case scenarios, a requirements document was compiled within each of the three use case categories. These three documents differed in structure and in the vocabulary used, and emphasised different aspects. For example, while the Data Annotation use case emphasised the need for a rich set of metadata descriptors, the Emotion Recognition use case pointed out the need to refer to sensor data, the use case on Emotion Generation requested a representation for the "reward" vs. "penalty" value of things. The situation was complicated further by the use of system-centric concepts such as "input" and "output", which for Emotion Recognition have fundamentally different meanings than for Emotion Generation.

In order to allow for an integration of the three requirements documents into one, 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".

Based on these principles and a large number of smaller clarifications, the three use case specific requirements documents were merged into an integrated wiki document[3]. After several iterations of restructuring and refinement, a consolidated structure has materialised for that document; in the following, we report on the key aspects.

3.1 Core emotion description

The most difficult aspect of the entire enterprise of proposing a generic emotion markup is the question of how to represent emotions. Given the fact that even emotion theorists have very diverse definitions of what an emotion is, and that very different representations have been proposed in different research strands (see e.g. [13] for an overview), any attempt to propose a standard way of representing emotions for technological contexts seems doomed to failure.

_____
[3] http://www.w3.org/2005/Incubator/emotion/wiki/UseCasesRequirements

The only viable way seems to be to give users a choice. Rather than trying to impose any of the existing emotion descriptions as the "correct" representation, the markup should provide the user with a choice of representations, so that an adequate representation can be used for a given application scenario.

This kind of choice should start with the possibility to explicitly state which type of affective or *emotion-related state* is actually being annotated. Different lists of such states have been proposed; for example, Scherer [14] distinguishes emotions, moods, interpersonal stances, preferences/attitudes, and affect dispositions.

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. We do not intend to impose any fixed set of emotion categories, because the appropriate categories will depend on the application. However, we can draw on existing work to propose a recommended set of emotion categories, which can be used if there are no reasons to prefer a different set. For example, [4] proposes a structured list of 48 emotion words as a candidate for a standard list.

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. The most well-known scales, sometimes by different names, are valence, arousal and potency; a recent large-scale study suggests that a more appropriate list may be valence, potency, arousal, and unpredictability [15]. Again, rather than imposing any given set of dimensions, the markup should leave the choice to the user, while proposing a recommended set that can be used by default.

As a third way to characterise emotions and related states, *appraisal scales* can be used, which provide details of the individual's evaluation of his/her environment. Examples include novelty, goal significance, or compatibility with one's standards. Again, a recommended set of appraisals may follow proposals from the literature (e.g., [16]), while the user should have the choice of using an application-specific set.

An important requirement for all three use cases was the fact that it should be possible to represent *multiple* and *complex emotions*. Different types of co-presence of emotions are envisaged: simultaneous emotions experienced due to the presence of several triggers (such as being sad and angry at the same time, but for different reasons); and regulation (such as trying to mask one emotion with another one, see below).

Emotions can have an *intensity*.

The concept of *regulation* [17] covers various aspects of an individual's attempts to feel or express something else than an emotion that spontaneously arises. On the behaviour level, that can lead to a difference between the "internal" and the "externalised" state. The various kinds of regulation which can be

envisaged include: *masking* one state with another one; *simulating* a state which is not present; and *amplifying* or *attenuating* a state.

Finally, it is required that some *temporal aspects* of the emotion be represented, including a start time and duration, and possibly changes of intensity or scale values over time.

3.2   Meta information about emotion description

Three additional requirements with respect to meta information have been elaborated: information concerning the degree of acting of emotional displays, information related to confidences and probabilities of emotional annotations, and finally the modalities involved. All of this information thereby applies to each annotated emotion separately.

Acting, which is particularly relevant for the Database Annotation use case, needs to cover the degree of naturalness, authenticity, and quality of an actor's portrayal of emotions, as e.g. perceived by test-subjects or annotators (an example of a database providing such information is [18]). In general, such attributes may be naturally quantified by use of a scale ranging from 0 to 1, to reflect for example the mean judgement among several test subjects or labellers.

Confidences and probabilities may generally be of interest for any of the three general use cases of annotation, recognition and synthesis. In the case of recognition, these are of particular importance within the multimodal integration of several input cues to preserve utmost information for a final decision process. Otherwise, a system reacting on emotions should be provided with additional information regarding the certainty of an assumed emotion to optimise the reaction strategy. In the case of database annotation, the mean inter-labeller agreement may be named as a typical example. More generally, it should be allowed to add such information to each level of representation, such as categories, dimensions, intensity, regulation, or degree of acting. Similar to the aforementioned meta information, confidences and probabilities may be represented by continuous scales, which preserves more information in a fusion scenario, or represented by symbolic labels as extra-low, low, medium, etc., which will often suffice to decide on a reaction strategy, e.g. in a dialogue.

The *modality* in which the emotion is reflected – observed or generated – is another example of a set that has to be left open for future additions. Typical generic modalities on a higher level are face, voice, body, text, or physiological signals; these can of course be further differentiated: parts of the face or body, intonation, text colour – the list of potential domain specific modalities is endless. Therefore, a core set of generally available modalities needs to be distinguished from an extensible set of application-specific modalities.

3.3   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.

One type of link which is required is a method for *linking to external media* objects, such as a text file containing the words of an utterance, an audio file, a video file, a file containing sensor data, technical description of sensor specifics, data enhancements applied, etc. This may for example be realised by a URL in an XML node.

A second kind of link deals with temporal linking to a *position on a time-line*. More specifically, this can be start and end times in absolute terms, or relative timings in relation to key landmarks on the time axis.

A mechanism should be defined for flexibly assigning meaning to those links. We identified the following initial set of meanings for such links to the "rest of the world": 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". Note that trigger and target are conceptually different; they may or may not coincide. As an illustration, consider the example of someone incidentally spilling coffee on one's clothing: though the trigger might be the cloth-ruining event, the target would be the person spilling the coffee.

We currently think that the links to media are relevant for all these semantics. Timing information seems to be relevant only for the observable behaviour and the trigger of an emotion.

3.4 Global metadata

Representing emotion, would it be for annotation, detection or generation, requires the description of the context not directly related to the description of emotion per se (e.g. the emotion-eliciting event) but also the description of a *more global context* which is required for exploiting the representation of the emotion in a given application. Specifications of metadata for multimodal corpora have already been proposed in the ISLE Metadata Initiative; but they did not target emotional data and were focused on an annotation scenario. The joint specification of our three use cases led to the identification of the following features required for the description of this global context.

For *person(s)*, we identified the following information as being potentially relevant: ID, date of birth, gender, language, personality traits (e.g. collected via personality questionnaires such as EPI for the annotation use case), culture, level of expertise as labeller. These pieces of information can be provided for real persons as well as for computer-driven agents such as ECAs or robots. For example, in the Data Annotation use case, it can be used for providing information about the subjects as well as the labellers.

Information about the *intended application* was also pointed out as being relevant for the exploitation of the representations of emotion (e.g. purpose of classification; application type – call centre data, online game, etc.; possibly, application name and version).

Furthermore, it should be possible to specify the *technical environment*. Within the document, it should be possible to link to that specification: for

example, the modality tag could link 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. For Data Annotation, this includes the type of collected data: fiction (movies, theatre), in-lab recording, induction, human-human interactions, human-computer interaction (real or simulated). All use cases might need the representation of metadata about the situational context in which an interaction occurs (number of people, relations, link to description of individual participants). Such information is likely to be global to an entire emotion markup document. It will be up to the application to use these in a meaningful way.

4   Conclusion and Outlook

In this paper, we have presented a consolidated list of requirements for a widely usable emotion markup language, based on a rich collection of use cases from a broad range of domains. This list aims at a balance between the aim of genericity and the fact that very different representations are required in different contexts. We are certain that the current list is not perfect; indeed, it is quite probably that we have missed out on some very relevant aspects. Despite these reservations, we believe that we have made reasonable progress towards a comprehensive list of requirements which can ultimately lead to a standard representation.

The next step will be to evaluate existing markup languages with respect to these requirements, in order to take stock of existing solutions for our needs. We also intend to sketch possible syntactic realisations of some of the key elements of the language.

Given the fact that the Emotion Incubator group is drawing to a close, serious work on a syntactic realisation will not be started within the lifetime of the group. Key design issues, such as the choice between XML and RDF formats, or the guiding principles of simplicity vs. non-ambiguity, deserve careful thinking. We are currently investigating possibilities for a follow-up activity, where an actual markup specification can be prepared.

Acknowledgements

References

1. Prendinger, H., Ishizuka, M.: Life-like Characters. Tools, Affective Functions and Applications. Springer (2004)
2. Ekman, P.: Facial expression and emotion. American Psychologist **48** (1993) 384–392

3. Schröder, M., Pirker, H., Lamolle, M.: 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 (2006) 88–92
4. Douglas-Cowie, E., et al.: HUMAINE deliverable D5g: Mid Term Report on Database Exemplar Progress. http://emotion-research.net/deliverables (2006)
5. Ekman, P., Friesen, W.: The Facial Action Coding System. Consulting Psychologists Press, San Francisco (1978)
6. Tekalp, M., Ostermann, J.: Face and 2-d mesh animation in MPEG-4. Image Communication Journal **15** (2000) 387–421
7. Devillers, L., Vidrascu, L., Lamel, L.: Challenges in real-life emotion annotation and machine learning based detection. Neural Networks **18** (2005) 407–422
8. Bevacqua, E., Raouzaiou, A., Peters, C., Caridakis, G., Karpouzis, K., Pelachaud, C., Mancini, M.: Multimodal sensing, interpretation and copying of movements by a virtual agent. In: Proceedings of Perception and Interactive Technologies (PIT'06). (2006)
9. Batliner, A., et al.: Combining efforts for improving automatic classification of emotional user states. In: Proceedings IS-LTC 2006. (2006)
10. Blech, M., Peter, C., Stahl, R., Voskamp, J., Urban, B.: Setting up a multimodal database for multi-study emotion research in HCI. In: Proceedings of the 2005 HCI International Conference, Las Vegas (2005)
11. Peter, C., Herbon, A.: Emotion representation and physiology assignments in digital systems. Interacting With Computers **18** (2006) 139–170
12. Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thórisson, K., Vilhjálmsson, H.: 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 (2006) 205–217
13. Cornelius, R.R.: The Science of Emotion. Research and Tradition in the Psychology of Emotion. Prentice-Hall, Upper Saddle River, NJ (1996)
14. Scherer, K.R.: Psychological models of emotion. In Borod, J.C., ed.: The Neuropsychology of Emotion. Oxford University Press, New York (2000) 137–162
15. Roesch, E., Fontaine, J., Scherer, K.: The world of emotion is two-dimensional – or is it? Presentation at the HUMAINE Summer school, Genoa, Italy (2006)
16. Scherer, K.R.: On the nature and function of emotion: A component process approach. In Scherer, K.R., Ekman, P., eds.: Approaches to emotion. Erlbaum, Hillsdale, NJ (1984) 293–317
17. Gross, J.J., ed.: Handbook of Emotion Regulation. Guilford Publications (2006)
18. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., Weiss, B.: A database of german emotional speech. In: Proc. Interspeech 2005, Lisbon, Portugal, ISCA (2005) 1517–1520