

# Chapter 4

## EmotionML

**Felix Burkhardt, Catherine Pelachaud, Björn W. Schuller,  
and Enrico Zovato**

**Abstract** EmotionML is a W3C recommendation to represent emotion related states in data processing systems. Given the lack of agreement in the literature on the most relevant aspects of emotion, it is important to provide a relatively rich set of descriptive mechanisms. It is possible to use EmotionML both as a standalone markup and as a plug-in annotation in different contexts. Emotions can be represented in terms of four types of descriptions taken from the scientific literature: categories, dimensions, appraisals, and action tendencies, with a single `<emotion>` element containing one or more of such descriptors. EmotionML provides a set of emotion vocabularies taken from the scientific and psychology literature. Whenever users have a need for a different vocabulary, however, they can simply define their own custom vocabulary and use it in the same way as the suggested vocabularies. Several applications have already been realized on the basis of EmotionML.

### 4.1 Introduction

EmotionML is a W3C recommendation to represent emotion related states in data processing systems. It was developed by a subgroup of the W3C MMI (Multimodal Interaction) Working Group chaired by Deborah Dahl in a first version from

---

F. Burkhardt (✉)

Telekom Innovation Laboratories, Winterfeldstr. 21, 10785 Berlin, Germany

e-mail: [Felix.Burkhardt@telekom.de](mailto:Felix.Burkhardt@telekom.de)

C. Pelachaud

LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, Paris, France

e-mail: [catherine.pelachaud@telecom-paristech.fr](mailto:catherine.pelachaud@telecom-paristech.fr)

B.W. Schuller

Imperial College, London, UK

University of Passau, Chair CIS, Passau, Germany

e-mail: [schuller@ieee.org](mailto:schuller@ieee.org)

E. Zovato

Nuance, Turin, Italy

e-mail: [ezovato@gmail.com](mailto:ezovato@gmail.com)

approximately 2005 until 2013, most of this time the development was led by Marc Schröder.

In the scientific literature on emotion research, there is no single agreed description of emotions, not even a clear consensus on the use of terms like affect, emotion, or other related phenomena. For a markup language representing emotional phenomena it therefore appears important to allow the representation of their most relevant aspects in the wider sense. Given the lack of agreement in the literature on the most relevant aspects of emotion, it is inevitable to provide a relatively rich set of descriptive mechanisms.

It is possible to use EmotionML both as a standalone markup and as a plug-in annotation in different contexts. Emotions can be represented in terms of four types of descriptions taken from the scientific literature: categories, dimensions, appraisals, and action tendencies, with a single `<emotion>` element containing one or more of such descriptors.

This chapter starts with a report on the analysis of use cases and requirements in Sect. 4.2. Section 4.3 gives a brief overview on previous attempts to formalize emotional vocabularies. In the following we discuss the elements that constitute EmotionML in Sect. 4.4. Section 4.5 motivates the suggested emotion vocabularies that were published as a W3C Note. Lastly, Sect. 4.6 introduces applications that were realized on the basis of EmotionML. The chapter closes with conclusions and outlook in Sect. 4.7.

## 4.2 Use Cases and Requirements

Although there are differences in emotion modeling theories, there is a sort of consensus on the essential components of emergent emotions. Among these the most important are feelings, appraisal, action tendencies, and emotion expressions (for example expressed in facial expressions, body gestures, and vocal cues). A shared formalism of encoding these aspects of emotions is a key point to allow inter-operability in technological environments.

The goal of representing emotions in a markup language was addressed with a bottom-up approach. Use cases were gathered from contributors with different expertise in the field of emotion in technology and research. The working group iteratively extracted requirements on the markup language from a number of 39 collected use cases. Based on these requirements, a syntax for EmotionML has been produced.

The collected use cases can be classified into three broader categories:

- Data Annotation
- Emotion Recognition
- Emotion Generation

Data annotation class includes many use cases that deal with human annotation of the emotion contained in some material, for example by means of a simple text,

or a node in an XML tree, representing pictures or voice recordings. The main requirement is defining the scope of the annotated emotion together with the emotion annotation itself. Annotation of emotions can be an important factor in training data for emotion classification. In data annotation it is sometimes necessary to keep track of the evolution of the emotion over time. An important requirement is in fact the possibility to trace events over time, in order to annotate the “dynamic” aspects of emotions, beyond the static ones.

The Emotion Recognition use cases deal with the detection of low level features from human–human or human–machine interaction. These features could be, for example, speech prosodic features [1] or facial action parameters [2]. Recognition can be unimodal or multimodal, where individual modalities can be merged. In this case timing and synchronization mechanisms are required. In emotion recognition also confidence measures have to be taken into consideration.

The Emotion Generation use cases deal mainly with the generation of face and body gestures and the generation of emotional speech. Emotion eliciting events trigger the generation of emotional behavior, through a mapping mechanism in which certain states are associated with specific behavioral actions.

These use cases showed that there is a big variety of information that is passed to and received from an emotion processing system. As a consequence, the emotion markup language had among its requirements a flexible method of receiving and sending data. The definition of which kind of information can be handled had to be specified as well.

The analysis of use cases showed that one of the primary requirements for the emotion markup language was the definition of a way of interfacing with external representations of data involved in the processes. In other words this markup language should not try to represent the expression of emotion itself, for example facial expressions or sensor data. Other languages should be used for this purpose. Also the system oriented concepts of input and output are replaced by more specific concepts like “experiencer,” “trigger,” and “observable behavior.”

The emotion markup language has among its requirements the possibility to describe emotions both in terms of categories, defined in specific sets (see Sect. 4.5), and of dimensions, for example “evaluation,” “activation,” and “power.” Beyond categories and dimensions the emotion markup language includes the possibility to specify “action tendencies” which are tendencies linked to specific emotions (e.g., anger can be linked to a tendency to attack). Also emotion “appraisals” can be described. These are important in cases where emotions play a role in driving behavior, i.e. when emotions are used to model the behavior of a person.

EmotionML has also to deal with mixed emotions and has to include the possibility to define a value in a well-defined scale for each of the component emotions. Temporal aspects are also important. The emotion markup language has to specify absolute times, relative times, and interval durations. As previously mentioned an important requirement is the specification of the time evolution of a dynamic scale value, through a tracing mechanism.

Meta information has to be also taken into account, such as the specification of the degree of confidence or probability that the represented emotion is correct.

The modality through which an emotion is expressed is also useful information. A general mechanism to specify meta information has been added in EmotionML allowing the definition of key-value pairs, in a similar fashion as in the EMMA markup language (@todo: link to EMMA chapter).

A mechanism to link the emotion description to external source of data is mandatory, since many use cases rely on media representations like video or audio files, pictures, documents, etc. Also the semantics of the external link has to be specified, and EmotionML fulfills this requirement by means of a specific attribute (“role”), that indicates whether the referenced link points to an observable behavior expressing the emotion, to the subject experiencing the emotion, to an emotion eliciting event, or to an object towards which an emotion-related action is directed.

### 4.3 Previous Work

The representation of emotions and related states has been part of several activities. For example, as part of their Virtual Human Markup Language, the Curtin University of Technology introduced in the Interface project EML, Emotion Markup Language, a vocabulary used to control the emotional expression of virtual characters.

In the area of labelling schemes, maybe the most thorough attempt to propose an encompassing labelling scheme for emotion-related phenomena has been the work on the HUMAINE database [3].

The relevant concepts were identified in prose, and made available as a set of configuration files for the video annotation tool Anvil [4]. A formal representation format was not proposed in this work. Markup languages including emotion-related information were defined mainly in the context of research systems generating emotion-related behaviour of ECAs (Embodied Conversational Agent).

The expressive richness is usually limited to a small set of emotion categories, possibly an intensity dimension, and in some cases a three-dimensional continuous representation of activation-evaluation-power space (see [5] for a review).

For example, the Affective Presentation Markup Language APML [6] provides an attribute “affect” to encode an emotion category for an utterance (a “performative”) or for a part of it:

```
<performative type="inform" affect="afraid">
Do I have to go to the dentist?
</performative>
```

An interesting contribution to the domain of computerized processing and representation of emotion-related concepts is A Layered Model of Affect, ALMA [7]. Following the OCC model [8], ALMA uses appraisal mechanisms to trigger emotions from events, objects, and actions in the world. Emotions have an intensity varying over time. Each individual emotion influences mood as a longer-term

affective state. ALMA uses an XML-based markup language named AffectML in two places: to represent the antecedents to emotion, i.e. the appraisals leading to emotions, or to represent the impact that emotions and moods have on a virtual agent's behaviour.

## 4.4 Emotion Markup Language Elements

Based on the requirements, a syntax for EmotionML has been produced in a sequence of steps.

The following snippet exemplifies the principles of the EmotionML syntax.

```
<sentence id="sent1">
  Do I have to go to the dentist?
</sentence>
<emotion xmlns="http://www.w3.org/2009/10/emotionml"
category-set="http://.../xml#everyday-categories">
  <category name="afraid" value="0.4"/>
  <reference role="expressedBy" uri="#sent1"/>
</emotion>
```

The following properties can be observed:

- The emotion annotation is self-contained within an `<emotion>` element.
- All emotion elements belong to a specific namespace.
- It is explicit in the example that emotion is represented in terms of categories.
- It is explicit from which category set the category label is chosen.
- The link to the annotated material is realized via a reference.

EmotionML is conceived as a plug-in language, with the aim to be usable in many different contexts. Therefore, proper encapsulation is essential. All information concerning an individual emotion annotation is contained within a single `<emotion>` element. All emotion markup belongs to a unique XML namespace.

EmotionML differs from many other markup languages in the sense that it does not enclose the annotated material. In order to link the emotion markup with the annotated material, either the reference mechanism in EmotionML or another mechanism external to EmotionML can be used.

A top-level element `emotionml` enables the creation of stand-alone EmotionML documents, essentially grouping a number of emotion annotations together, but also providing document-level mechanisms for annotating global meta data and for defining emotion vocabularies (see below). It is thus possible to use EmotionML both as a standalone markup and as a plug-in annotation in different contexts.

### 4.4.1 *Representations of Emotion*

Emotions can be represented in terms of four types of descriptions taken from the scientific literature [9]: `<category>`, `<dimension>`, `<appraisal>`, and `<action-tendency>`. An `<emotion>` element can contain one or more of these descriptors; each descriptor must have a `name` attribute and can have a `value` attribute indicating the intensity of the respective descriptor. For `<dimension>`, the `value` attribute is mandatory, since a dimensional emotion description is always a position on one or more scales; for the other descriptions, it is possible to omit the `value` to only make a binary statement about the presence of a given category, appraisal or action tendency.

The following example illustrates a number of possible uses of the core emotion representations.

```
<category name="affectionate" />
<dimension name="valence" value="0.9" />
<appraisal name="agent-self" />
<action-tendency name="approach" />
```

### 4.4.2 *Mechanism for Referring to an Emotion Vocabulary*

Since there is no single agreed-upon vocabulary for each of the four types of emotion descriptions, EmotionML provides a mandatory mechanism for identifying the vocabulary used in a given `<emotion>`. The mechanism consists in attributes of `<emotion>` named `category-set`, `dimension-set`, etc., indicating which vocabulary of descriptors for annotating categories, dimensions, appraisals, and action tendencies are used in that emotion annotation. These attributes contain a URI pointing to an XML representation of a vocabulary definition. In order to verify that an emotion annotation is valid, an EmotionML processor must retrieve the vocabulary definition and check that every `name` of a corresponding descriptor is part of that vocabulary.

Some vocabularies are suggested by the W3C [10]. Users are encouraged to use them to make EmotionML documents interoperable.

### 4.4.3 *Meta-Information*

Several types of meta-information can be represented in EmotionML.

First, each emotion descriptor (such as `<category>`) can have a `confidence` attribute to indicate the expected reliability of this piece of the annotation. This can

reflect the confidence of a human annotator or the probability computed by a machine classifier. If several descriptors are used jointly within an `<emotion>`, each descriptor has its own `confidence` attribute. For example, it is possible to have high confidence in, say, the arousal dimension but be uncertain about the pleasure dimension:

```
<emotion dimension-set="http://www.w3.org/TR/emotion-voc/xml#pad-dimensions">
  <dimension name="arousal" value="0.7" confidence="0.9"/>
  <dimension name="pleasure" value="0.6" confidence="0.3"/>
</emotion>
```

Each `<emotion>` can have an `expressed-through` attribute providing a list of modalities through which the emotion is expressed. Given the open-ended application domains for EmotionML, it is naturally difficult to provide a complete list of relevant modalities. The solution provided in EmotionML is to propose a list of human-centric modalities, such as `gaze`, `face`, `voice`, etc., and to allow arbitrary additional values. The following example represents a case where an emotion is recognized from, or to be generated in, face and voice:

```
<emotion category-set="http://.../xml
# everyday-categories" expressed-through="face voice">
  <category name="satisfaction"/>
</emotion>
```

For arbitrary additional meta data, EmotionML provides an `<info>` element which can contain arbitrary XML structures. The `<info>` element can occur as a child of `<emotion>` to provide local meta data, i.e. additional information about the specific emotion annotation; it can also occur in standalone EmotionML documents as a child of the root node `<emotionml>` to provide global meta data, i.e. information that is constant for all emotion annotations in the document. This can include information about sensor settings, annotator identities, situational context, etc.

#### 4.4.4 References to the “Rest of the World”

Emotion annotation is always *about* something. There is a subject “experiencing” (or simulating) the emotion. This can be a human, a virtual agent, a robot, etc. There is observable behavior expressing the emotion, such as facial expressions, gestures, or vocal effects. With suitable measurement tools, this can also include physiological changes such as sweating or a change in heart rate or blood pressure. Emotions are often caused or triggered by an identifiable entity, such as a person, an object, an event, etc. More precisely, the appraisals leading to the emotion are triggered by

that entity. And finally, emotions, or more precisely the emotion-related action tendencies, may be directed towards an entity, such as a person or an object.

EmotionML considers all of these external entities to be out of scope of the language itself; however, it provides a generic mechanism for referring to such entities. Each `<emotion>` can use one or more `<reference>` elements to point to arbitrary URIs. A `<reference>` has a `role` attribute, which can have one of the following four values: `expressedBy` (default), `experiencedBy`, `triggeredBy`, and `targetedAt`. Using this mechanism, it is possible to point to arbitrary entities filling the above-mentioned four roles; all that is required is that these entities be identified by a URI.

#### 4.4.5 *Time*

Time is relevant to EmotionML in the sense that it is necessary to represent the time during which an emotion annotation is applicable. In this sense, temporal specification complements the above-mentioned reference mechanism.

Representing time is an astonishingly complex issue. A number of different mechanisms are required to cover the range of possible use cases. First, it may be necessary to link to a time span in media, such as video or audio recordings. For this purpose, the `<reference role="expressedBy">` mechanism can use a so-called Media Fragment URI to point to a time span within the media [11]. Second, time may be represented on an absolute or relative scale. Absolute time is represented in milliseconds since 1 January 1970, using the attributes `start`, `end`, and `duration`. Absolute times are useful for applications such as affective diaries, which record emotions throughout the day, and whose purpose is to link back emotions to the situations in which they were encountered. Other applications require relative time, for example time since the start of a session. Here, the mechanism borrowed from EMMA is the combination of `time-ref-uri` and `offset-to-start`. The former provides a reference to the entity defining the meaning of time 0; the latter is time, in milliseconds, since that moment.

#### 4.4.6 *Representing Continuous Values and Dynamic Changes*

As mentioned above, the emotion descriptors `<category>`, `<dimension>`, etc. can have a `value` attribute to indicate the position on a scale corresponding to the respective descriptor. In the case of a dimension, the value indicates the position on that dimension, which is mandatory information for dimensions; in the case of categories, appraisals, and action tendencies, the value can be optionally used to indicate the extent to which the respective item is present.



In all cases, the `value` attribute contains a floating-point number between 0 and 1. The two end points of that scale represent the most extreme possible values, for example the lowest and highest possible positions on a dimension, or the complete absence of an emotion category vs. the most intense possible state of that category.

The `value` attribute thus provides a fine-grained control of the position on a scale, which is constant throughout the temporal scope of the individual `<emotion>` annotation. It is also possible to represent changes over time of these scale values, using the `<trace>` element which can be a child of any `<category>`, `<dimension>`, `<appraisal>`, or `<action-tendency>` element. This makes it possible to encode trace-type annotations of emotions as produced.

## 4.5 Vocabularies

As described above, EmotionML takes into account a number of key concepts from scientific emotion research [5]. Four types of descriptions are available: categories, dimensions, appraisals, and action tendencies. These types correspond to the four main existing representation scheme of emotions.

Depending on the tradition of emotion research and on the use case, it may be appropriate to use any single one of these representations; alternatively, it may also make sense to use combinations of descriptions to characterize more fully the various aspects of an emotional state that are observed: how the value attributed to an appraisal caused the emotion; how an emotion can be described in terms of a category and/or a set of dimensions; and the potential actions an individual may be executing when an emotion is triggered. Insofar, EmotionML is a powerful representational device.

This description glosses over one important detail, however. Whereas emotion researchers may agree to some extent on the types of facets that play a role in the emotion process (such as appraisals, feeling, expression, etc.), there is no general consensus on the representation schema nor on the descriptive vocabularies that should be used. Which set of emotion categories is considered appropriate varies dramatically between the different traditions, and even within a tradition such as the Darwinian tradition of emotion research, there is no agreed set of emotion categories that should be considered as the most important ones (see, e.g., [12]). Similarly, emotion theoreticians do not agree on the number nor the type of dimensions to consider. Similar remarks can be made for appraisals.

For this reason, any attempt to enforce a closed set of descriptors for emotions would invariably draw heavy criticism from a range of research fields. Given that there is no consensus in the community, it is impossible to produce a consensus annotation in a standard markup language. The obvious alternative is to leave the choice of descriptors up to the users; however, this would dramatically limit interoperability.

The solution pursued in EmotionML is of a third kind. The notion of an ‘emotion vocabulary’ is introduced: any specific emotion annotation must be specific about the vocabulary that is being used in that annotation. This makes it possible to define in a clear way the terms that make sense in a given research tradition. Components that want to interoperate need to settle on the emotion vocabularies to use; whether a given piece of EmotionML markup can be meaningfully interpreted by an EmotionML engine can be determined.

The specification includes a mechanism for defining emotion vocabularies. It consists of a ‘<vocabulary>’ element containing a number of ‘<item>’ elements. A vocabulary has a ‘type’ attribute, indicating whether it is a vocabulary for representing categories, dimensions, appraisals, or action tendencies. A vocabulary item has a ‘name’ attribute. Both the entire vocabulary and each individual item can have an ‘<info>’ child to provide arbitrary metadata.

A W3C Working Group Note [13] complements the specification to provide EmotionML with a set of emotion vocabularies taken from the scientific and psychology literature. When the user considers them suitable, these vocabularies rather than other arbitrary vocabularies should be used in order to promote interoperability.

Whenever users have a need for a different vocabulary, however, they can simply define their own custom vocabulary and use it in the same way as the vocabularies listed in the Note. This makes it possible to add any vocabularies from scientific research that are missing from the pre-defined set, as well as application-specific vocabularies.

In selecting emotion vocabularies, the group has applied the following criteria. The primary guiding principle has been to select vocabularies that are either commonly used in technological contexts, or represent current emotion models from the scientific and psychology literature. A further criterion is related to the difficulty to define mappings between categories, dimensions, appraisals, and action tendencies.

For this reason, groups of vocabularies were included for which some of these mappings are likely to be definable in the future.

The following vocabularies are defined. For categorical descriptions, the “big six” basic emotion (often referred to as universal) vocabulary by Ekman [4], an everyday emotion vocabulary by Cowie et al. [14], and three sets of categories that lend themselves to mappings to, respectively, appraisals, dimensions, and action tendencies: the 22 OCC labels of emotion categories [8], the 24 labels used by Fontaine et al. [15] that are further defined in terms of 4 dimensions, and the 12 categories that are linked to actions tendency as introduced by Frijda [16].

Three dimensional vocabularies are provided, the pleasure-arousal-dominance (PAD) vocabulary by Mehrabian [17], the four-dimensional vocabulary proposed by Fontaine et al. [15], and a vocabulary providing a single ‘intensity’ dimension for such use cases that want to represent solely the intensity of an emotion without any statement regarding the nature of that emotion. For appraisal, three vocabularies are proposed: the appraisals defined in the OCC model of emotions [8], Scherer’s Stimulus Evaluation Checks which are part of the Component Process

Model of emotions [18], and the appraisals used in the computational model of emotions EMA [19]. Finally, for action tendencies, only a single vocabulary is currently listed, namely that proposed by Frijda [10]. The following example represents the call for an emotion described by its four dimensions based on Fontaine et al. [16]:

```
<emotion dimension-set="http://www.w3.org/TR/emotion-voc/xml
  #fsre -dimensions">
<dimension name = "valence " value="0.8"/> <! high positive
valence -->
<dimension name = "arousal " value="0.5"/> <! average arousal -->
<dimension name = "potency" value="0.2"/> <! very low potency -->
<dimension name = "unpredictability " value="0.4"/> <! some-
how lower than average predictable event -->
</emotion>
```

While these vocabularies should provide users with a solid basis, it is likely that additional vocabularies or clarifications about the current vocabularies will be requested. Due to the rather informal nature of a W3C Note, it is rather easy to provide future versions of the document that provide the additional information required.

## 4.6 Applications

In the following, a range of applications is named that provide EmotionML support on the input and/or output side. While a complete listing is not possible due to the increasing usage of the standard, a selection has been made on the criteria to (a) cover for the different use-cases as listed above and (b) based on popularity and spread of the solutions to solve these. Some actually delivered implementation reports to the recommendation.<sup>1</sup> The applications are grouped.

### 4.6.1 Data Annotation

GTrace provides a tool for the annotation of continuous dimensional emotion in the sense of “traces” whether the emotion primitive such as arousal or valence is rising or falling over time [20]. This is done via a mouse (or joystick, etc.) in a 1D window that is shown side-by-side with the material to be annotated. Various pre-specified

---

<sup>1</sup><https://www.w3.org/2002/mmi/2013/emotionml-ir/>.

scales are provided and one can customize these or add new ones. The program is fully compatible with EmotionML.

The Speechalyzer by Deutsche Telekom Laboratories is an open source project for analysis, annotation, and transcription of speech files [21]. It can be used to rapidly judge large numbers of audio files emotionally, an automatic classification is integrated. The Speechalyzer was part of a project to identify disgruntled customers in an automated voice service portal [22] with two use cases in mode: (a) transfer angry users to a trained human agent, and (b) gain some statistic insight on the number of angry customers at the end of each day. It utilizes EmotionML as an exchange format to import and export emotionally annotated speech data.

iHEARu-PLAY is a gamified crowd-sourcing platform [23] that supports audio, image, and video annotation for emotion supporting EmotionML. It also allows for crowd-sourced recording of data. At present, dynamic active learning abilities are integrated.

Further examples of annotation software supporting EmotionML include DocEmoX, a tool to annotate documents in 3D emotion space [24],

## 4.6.2 *Emotion Recognition*

The openSMILE tool first developed during the European SEMAINE project supports extraction of large audio feature spaces in real time for emotion analysis from audio and video [25] and has also been used for other modalities such as physiological data or CAN-Bus data in the car. It is written in C++ and has been ported to Android for mobile usage. The main features are the capability of on-line incremental processing and high modularity that allows for feature extractor components to be freely interconnected for the creation of custom features via a simple configuration file. Further, new components can be added to openSMILE via an easy plugin interface and a comprehensive API. openSMILE is free software licensed under the GPL license. The toolkit has matured to a standard in the field of Affective Computing also due to its usage in a broad range of competitions in the field including AVEC 2011–2016, Interspeech ComParE 2009–2016, EmotiW, and MediaEval.

The openEAR extension of openSMILE provides pre-trained models for emotion recognition and a ready-to-use speech emotion recognition engine [26].

The EyesWeb platform was enhanced in the ASC-Inclusion European project enabling it to send text messages containing emotion Markup Language messages to give information about recognized emotions from body gestures [27]. EyesWeb is an open platform that supports a wide number of input devices including motion capture systems, various types of professional and low cost video cameras, game interfaces (e.g., Kinect, Wii), multichannel audio input (e.g., microphones), and analog inputs (e.g., for physiological signals). Supported outputs include multichannel audio, video, analog devices, and robotic platforms.

### 4.6.3 *Emotion Generation*

MARY TTS is an open-source, multilingual text-to-speech synthesis platform that includes modules for expressive speech synthesis [28]. Particularly the support for both categorical and dimensional representations of emotions by EmotionML is important to Mary's expressive speech synthesis. These categories and dimensions are implemented by modifying the predicted pitch contours, pitch level, and speaking rate.

Using this approach, expressive synthesis is most effective when using HMM-based voices, since the statistical parametric synthesis framework allows appropriate prosody to be realized with consistent quality. Expressive unit selection voices support EmotionML best if they are built from multiple-style speech databases [29], which preserve intonation and voice quality better than when applying signal manipulation to conventional unit-selection output.

Greta is a real-time 3D embodied conversational agent with a 3D model of an agent compliant with MPEG-4 animation standard [30]. It is able to communicate using a rich palette of verbal and nonverbal behaviours. Greta can talk and simultaneously show facial expressions, gestures, gaze, and head movements. Besides the standard XML languages FML and BML that allow to define the communicative intentions and behaviours, EmotionML support was added and used in a range of European projects such as SEMAINE, TARDIS, and ARIA-VALUSPA.

### 4.6.4 *Platforms and Projects*

The SEMAINE platform [31] stems from the European Semaine-Project. It provides a free to use virtual agent system including full audio/visual input analysis (e.g., via openSMILE) and output generation (via MARY TTS and Greta) as well as a dialogue manager. The communication between modules is based on EmotionML.

Finally, a range of projects use the standard. Examples are the above named SEMAINE and ARIA-VALUSPA European projects, both dealing with audiovisual emotionally intelligent chatbots. Further the ASC-Inclusion and De-ENIGMA projects aiming to help children on the autism spectrum to learn about emotions, the TARDIS European project that provides serious gaming to young individuals to prepare for their first job interviews, the MixedEmotions and SEWA European projects focusing on sentiment analysis as well as a range of national projects such as the Finnish "Detecting and visualizing emotions and their changes in text" project [32, 33].

## 4.7 Conclusions

We presented EmotionML, a W3C recommendation to represent emotion related states in data processing systems.

It is possible to use EmotionML both as a standalone markup and as a plug-in annotation in different contexts. Emotions can be represented in terms of four types of descriptions taken from the scientific literature: categories, dimensions, appraisals, and action tendencies, with a single `<emotion>` element containing one or more of such descriptors.

A W3C Working Group Note complements the specification to provide EmotionML with a set of suggested emotion vocabularies taken from the scientific and psychology literature. Whenever users have a need for a different vocabulary, however, they can simply define their own custom vocabulary and use it in the same way as the vocabularies listed in the Note.

Several applications have already been realized on the basis of EmotionML but we're still far away from widespread use. This of course reflects the fact that emotion processing systems are still in their technological infancy and up to now are more research topic than product feature.

Nonetheless we believe that with the spreading of user interfaces that are more natural than keyboard typing, as, for example, speech interfaces, wearables or physical sensors, emotional processing will become a necessity for such systems to be able to interact in a natural and intuitive manner.

Another technology trend that pushes emotional processing is the renaissance of artificial intelligence, as intelligence and emotions are strongly connected concepts.

We hope this article encourages the reader to use EmotionML in her/his own projects and give feedback to the W3C to pave the way towards EmotionML version 2.0. Some topics that came up in the group's discussions have been left out in the first version, for the sake of simplicity. For example, the blend of emotions, emotion regulation, or a direct link to RDF (Resource Description Framework) for semantic annotation was dropped at some point. Another issue that could be pursued are requirements that resolve from use cases concerned with sentiment analysis, which is an important topic given the automatic analysis of user generated content.

## References

1. Devillers, L., Vidrascu, L., & Lamel, L. (2005). Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks*, 18(4), 407–422 (2005 special issue).
2. Tekalp, A. M., & Ostermann, J. (2000). Face and 2-D mesh animation in MPEG-4. *Image Communication Journal*, 15, 387–421.
3. Douglas-Cowie, E., Cowie, R., Sneddon, I., Cox, C., Lowry, O., McRorie, M., et al. (2007). The HUMAINE database: Addressing the collection and annotation of naturalistic and induced

- emotional data. In *Proceedings of Affective Computing and Intelligent Interaction*, Lisbon, Portugal (pp. 488–500).
4. Kipp, M. (2014). ANVIL: a universal video research tool. In J. Durand, U. Gut, & G. Kristofferson (Eds.), *Handbook of corpus phonology*, pp. 420–436. Oxford: Oxford University Press.
  5. Schröder, M., Pirker, H., Lamolle, M., Burkhardt, F., Peter, C., & Zovato, E. (2011). Representing emotions and related states in technological systems. In P. Petta, R. Cowie, & C. Pelachaud (Eds.), *Emotion-oriented systems – The humane handbook* (pp. 367–386). Berlin: Springer.
  6. de Carolis, B., Pelachaud, C., Poggi, I., & Steedman, M. (2004). APLM, a markup language for believable behavior generation. In H. Prendinger & M. Ishizuka (Eds.), *Life-like characters* (pp. 65–85). New York: Springer.
  7. Gebhard, P. (2005). ALMA - A layered model of affect. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-05)*, Utrecht.
  8. Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotion*. Cambridge, UK: Cambridge University Press.
  9. Schröder, M., Pirker, H., Lamolle, M., Burkhardt, F., Peter, C., & Zovato, E. (2011). Representing emotions and related states in technological systems. In *Emotion-oriented systems - The humane handbook* (pp. 367–386). Berlin: Springer.
  10. Frijda, N. H. (1986). *The emotions*. Cambridge, UK: Cambridge University Press.
  11. Troncy, R., Mannens, E., Pfeiffer, S., & van Deursen, D. (2012, March 15). Media fragments URI 1.0: W3c proposed recommendation.
  12. Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication*, 40(1–2), 5–32.
  13. Schröder, M., Pelachaud, C., Ashimura, K., Baggia, P., Burkhardt, F., Oltramari, A., et al. (2011). Vocabularies for emotionml. <http://www.w3.org/TR/emotion-voc/>
  14. Cowie, R., Douglas-Cowie, E., Appolloni, B., Taylor, J., Romano, A., & Fellenz, W. (1999). What a neural net needs to know about emotion words. In N. Mastorakis (Ed.), *Computational intelligence and applications* (pp. 109–114). Singapore: World Scientific & Engineering Society Press.
  15. Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science*, 18(12), 1050–1057.
  16. Frijda, N. H. (1986). *The emotions*. Cambridge, UK: Cambridge University Press.
  17. Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4), 261–292.
  18. Scherer, K. R. (1999). Appraisal theory. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition & emotion* (pp. 637–663). New York: Wiley.
  19. Gratch, J., & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4), 269–306.
  20. Cowie, R., Sawey, M., Doherty, C., Jaimovich, J., Fyans, C., & Stapleton, P. (2013). Gtrace: General trace program compatible with emotionml. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 709–710). New York: IEEE.
  21. Burkhardt, F. (2011). Speechalyzer: A software tool to process speech data. In *Proceedings of the ESSV, Elektronische Sprachsignalverarbeitung*.
  22. Burkhardt, F., Polzehl, T., Stegmann, J., Metzke, F., & Huber, R. (2009). Detecting real life anger. In *Proceedings ICASSP*, Taipei, Taiwan (Vol. 4).
  23. Hantke, S., Appel, T., Eyben, F., & Schuller, B. (2015). iHEARu-PLAY: Introducing a game for crowd sourced data collection for affective computing. In *Proceedings of 1st International Workshop on Automatic Sentiment Analysis in the Wild (WASA 2015)*, Xi'an, P.R. China (pp. 891–897). New York: IEEE.

24. Kouroupetroglou, G., Tsonos, D., & Vlahos, E. (2009). Docemox: A system for the typography-derived emotional annotation of documents. In *Universal Access in Human-Computer Interaction. Applications and Services* (pp. 550–558). New York: Springer.
25. Eyben, F., Weninger, F., Groß, F., & Schuller, B. (2013). Recent developments in openSMILE, the Munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM International Conference on Multimedia, MM 2013*, Barcelona, Spain (pp. 835–838). New York: ACM.
26. Eyben, F., Wöllmer, M., & Schuller, B. (2009, September). openEAR – Introducing the Munich open-source emotion and affect recognition toolkit. In *Proceedings 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*, Amsterdam, The Netherlands (Vol. I, pp. 576–581). HUMAINE Association. New York: IEEE.
27. Piana, S., Staglianò, A., Camurri, A., & Odone, F. (2013). A set of full-body movement features for emotion recognition to help children affected by autism spectrum condition. In *IDGEI International Workshop*.
28. Charfuelan, M., & Steiner, I. (2013). Expressive speech synthesis in mary tts using audiobook data and emotionml. In *Proceedings of Interspeech*.
29. Steiner, I., Schröder, M., & Klepp, A. (2013). The PAVOQUE corpus as a resource for analysis and synthesis of expressive speech. *Proceedings of Phonetik & Phonologie* (Vol. 9).
30. Bevacqua, E., Prepin, K., Niewiadomski, R., de Sevin, E., & Pelachaud, C. (2010). Greta: Towards an interactive conversational virtual companion. In *Artificial Companions in Society: Perspectives on the Present and Future* (pp. 143–156).
31. Schröder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., et al. (2012). Building autonomous sensitive artificial listeners. *IEEE Transactions on Affective Computing*, 3(2), 165–183.
32. Munezero, M., Kakkonen, T., & Montero, C. S. (2011). Towards automatic detection of antisocial behavior from texts. In *Sentiment analysis where AI meets psychology (SAAIP)* (p. 20).
33. Burkhardt, F., Becker-Asano, C., Begoli, E., Cowie, R., Fobe, G., & Gebhard, P. (2014). Application of emotionml. In *Proceedings of the 5th International Workshop on Emotion, Sentiment, Social Signals and Linked Open Data (ES3LOD)*.