Modelling multimodal context-aware affective interaction

Idoia Cearreta, Juan Miguel López, Nestor Garay-Vitoria

Laboratory of Human-Computer Interaction for Special Needs, University of the Basque Country
Manuel Lardizabal 1, E-20018 Donostia (Gipuzkoa)
icearreta001@ikasle.ehu.es, juanmi@si.ehu.es,
nestor.garay@ehu.es

Abstract. In this paper a generic approach to modelling context-aware emotions taking different theoretical models into account is presented. Developing generic computational models for modelling emotions has become necessary due to the great variety of theoretical models of emotions and implementation technologies which can be used in the design and implementation of affective systems. The main component in this approach is a domain ontology of context-aware emotions. Concepts in ontology are grouped into several global modules, representing different aspects related to emotion modelling. Proposed ontology enables description of emotions at different levels of abstraction while serving as a guide for flexible design of affective context-aware applications (such as affective synthesizers and recognizers) independently of the starting model and the final way of implementation.

Keywords: Models of emotions, domain ontology, application ontology, software engineering, context, multimodality, emotional cues

1 Introduction

Affective computing is an area that works on the detection and response to user's emotions [1]. Currently, there is a great variety of theoretical models of emotions that can frame the design of affective applications, and there are different technologies that can be used for their implementation. In addition, although there are many common properties, emotions are not universal: they are differently expressed in different cultures and languages, while many emotional properties are individual. There is rarely a one-size-fits-all solution for the growing variety of computer users and interactions [1]. Therefore, emotionally-aware applications should be designed in a flexible way if they are wanted to be used with a wider class of users.

One of the main aims in affective computing is to produce tools that can be easily personalized to each user. Personalization is necessary for more efficient interaction, and for fine tuning and better acceptation of developed systems. Each person reflects emotions in a personal way, so there is a need to properly adjust parameters to each one of them.

Nowadays interaction is becoming user-centred and computers are not always in the same location. As it can be noted, mobility of both user and computer has increased in last years and, in this situation, studying context has been found relevant to user interaction. A new multidisciplinary paradigm called Ambient Intelligence (AmI) is emerging. In AmI, technologies are deployed to make computers disappear in the background while human user moves into the foreground in complete control of the augmented environment [2].

Including affective computing paradigm within AmI seems to be an interesting approach. User gestures (both facial and corporal), speeches, words, etc., can be used to detect people's emotional state by reading multimodal sources. It can be mentioned miniaturization of computer-based devices is making possible the development of wearable computers that can help in recording parameters, initially obtained in an intrusive way, without disturbing people.

With the aim of solving presented problems, on the one hand a generic approach of a domain ontology is proposed. This domain ontology defines concepts of context-aware emotions taking into account different theoretical models. On the other hand, an example application ontology is presented. This application ontology contains all the necessary concepts to model specific applications; in this case, affective synthesizers or recognizers in speech. Therefore, these ontologies can serve as guide for flexible design of multimodal affective devices with independence of the starting model and the final way of implementation taking into account multidisciplinary [3].

In next section, a brief revision of related work is included. After that, the basic idea of this approach is made known. Then, the domain ontology of context-aware emotions is introduced. Next, the example application ontology for speech modelling is presented. In the end, conclusions and future works are shown.

2 Related work

Theories of emotions proposed by cognitive psychologist are a useful starting point. Although several theoretical models of emotions exist, the most commonly used are the dimensional [4], categorical [5] and appraisal [6] models of emotions. For practical reasons, categorical models of emotions have been more frequently used in affective computing. However, appraisal theoretical model is probably the most influential approach to emotion within psychology at present [6]. Appraisal theory offers a descriptive framework for emotion based on the way the person involved experiences the events, things, or people at the focus of the emotional state. Anyway, dimensional approach to emotion has been advocated by a number of theorists, such as [7]. Emotion dimensions are a simplified description of basic properties of emotional states [8]. The most frequently encountered emotion dimensions are *activation*, *evaluation* and *power*.

Apart from models, there are also some studies related to expression and detection of emotions. In this way, [4] proposed that three systems exist that would be implied in the expression of the emotions and that could serve like indicators to detect the emotion of the user: Verbal information (reports about perceived emotions described by users); Conductal information (facial and postural expressions and

speech paralinguistic parameters); and Psychophysiological answers (such as heart rate, galvanic skin response –GSR–, and electroencephalographic response).

In general, research has paid little attention to context in affective computing area [9]. Context is inescapably linked to modality and emotion is strongly multimodal in the sense that signs may appear in various different channels. However, not all types of sign tend to be available together, because context will affect signs that are relevant or accessible.

Several authors have given some context definitions. For example, [10] mention that context-aware applications look at the who's, where's, when's and what's of entities and use this information to determine why the situation is occurring. [11] present *AmbieSense* system, where User context is composed by five parts: Environment context (includes environmental data, such as topics related to the place where user is); Personal context (includes personal data, physiological and mental); Task context (describes what the persons or actors are doing in this User context); Social context (describes the social aspects of the current User context); and Spatio-temporal context (describes aspects of the User context relating to the time and spatial extent for the User context).

In the Human-Computer Interaction (HCI) field, different models and approaches have been defined in order to better capture relevant concepts from user interaction, some of them based on software engineering paradigms. Object Management Group's (OMG) Model-Driven Architecture (MDA) is a software design approach to system specification and interoperability based on the use of hierarchically organized formal models [12]. MDA is supported by one of the most popular modelling languages, Unified Modeling Language (UML) [13]. UML consists of a collection of semiformal graphical notations. These notations can be used to support different development methodologies [14]. The work of [15] lists a number of good reasons why UML is a promising notation for ontologies. Besides the possibility to use existing modelling tools, standardization provides a significant driving force for further progress because it codifies best practices, enables and encourages reuse, and facilitates interworking between complementary tools [16]. W3C is also promoting standarization by means of the W3C emotion incubator group [17].

A generic ontology, as the one presented in this paper, can be useful for the description of emotions according to the different systems of emotional expression and detection, elements which compose user context and theoretical models of emotions. The work of [18] for example, presents a novel approach to affective sensing, using a generic model of affective communication and a set of ontologies to assist in the analysis of concepts and to enhance the recognition process, taking the context and different emotion classification methods into account.

3 Ontology for description of context-aware emotions

Allowing flexible definitions of emotions at different levels of abstraction is the basic motivation of this solution. A first step has been to develop a conceptual ontology that allows such logical division of concepts. An ontology is usually defined as an explicit specification of an abstract, simplified view of a world to be represented, specifying

both the concepts inherent in this view as well as their interrelationships [19]. In this case, an ontology can provide a common and standardized language for sharing and reusing knowledge about emotions.

Although general nature of ontologies make them independent of a specific modelling language, this ontology has been applied to extend UML in order to make research more practical, and more accessible to ordinary software engineers. UML is a widely adopted standard that is familiar to many software practitioners, extensively taught in undergraduate courses, and supported by many books, training courses, and tools from different vendors. In conclusion, the main aim to use this language is to create a graphical representation to make ontologies easier to understand.

The essence of proposed solution is a domain ontology of context-aware emotions, which introduces concepts and mechanisms used in the affective computing domain. This ontology can be viewed as a metamodel that formally describes important affective computing concepts, and introduces concepts used to create models of concrete emotions. Therefore, this ontology collects information obtained of different channels or systems [4], providing the development of multimodal affective applications, taking the user context into account. Authors have based on references found in the literature to define this ontology.

Concepts in the domain ontology are grouped into seven global modules representing different aspects related to emotions (see Figure 1). In Figure 1, emotion-related modules are grouped within the central square as they were defined in [3].

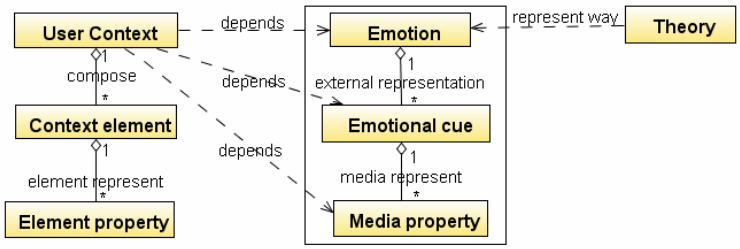

Fig. 1. Relationship among modules in the domain ontology of emotional concepts.

Emotion module. Defines the emotion of the user within a context. This emotion can be composed by one or more kinds of emotional cues (verbal, facial, gestural, speech and psychophysiological emotional cues [4]). The emotion is influenced by the context in which the user is and may changes throughout time for different reasons. For instance, taking the personal context into account, user can change from *joy* to *sadness* emotion when remembering a sad episode. Furthermore, emotions can be represented in different ways according to different theories. Therefore, different emotional values will be transmitted or collected through different emotional cues and depending on the context, it will produce or it will recognize the emotion of the user.

Theory module. Describes the main types of theories found in the literature, such as categorical [5], dimensional [4] and appraisal [6]. For each type of theory, either in categorical, dimensional or appraisal, the emotion will be represented in a

different way. It must be highlighted that although these three theories models have been mentioned, others theory models can also be used.

Emotional cue module. Describes external emotional representations in terms of different media properties. Depending on the context, an emotional cue will be taken into account more than another one.. For instance, in some cases sounds may not be allowed (as in the case of libraries). Each type of emotional cue corresponds to each one of the three systems proposed by [4] and taking into account all emotional cues, emotion is completed.

User context module. Defines the user context as composed by different context elements: environment, personal, task, social and spatio-temporal context [11]. Emotion is directly influenced by the context where user is in a concrete moment. It must be noticed that although in this work the user context has been described by [11] model, other context models can also be used.

Context element module. Describes the context representation in terms of the properties for the different elements. Each part of the user context which [11] suggested corresponds to a context element. There are some components like verbal cues which are related to the user language used to express himself/herself. All these factors have influence in emotion expression and identification. On the other hand, different emotional cues can be taken into account according to user context; for example, if it is dark, facial emotional cue may not be relevant, so speech emotional cue will be more relevant in this case.

Media property module. Describes basic media properties for emotional cues. These media properties are used for description of emotional cues and they are also context-aware. For example, taking personal context element into account, voice intensity value is different depending on people's gender. A media property can be basic, such as voice intensity, or derived, such as voice intensity variations (see Figure 2).

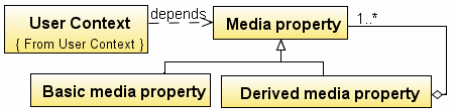

Fig. 2. Model of media depending on user context.

Element property module. Describes properties for context elements. An element property can be basic, such as the temperature of the environment, or derived, such as the mean temperature (see Figure 3). These element properties are used for description of context elements. For example, an environment context element can be composed by average temperature, average noise and light intensity, and the composition with the other context elements, user context is completed.

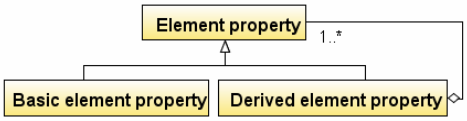

Fig. 3. Model of element property

4 An example application ontology for describing speech emotional cue

For defining a practical example using the proposed domain ontology, some modelling language must be chosen. UML [13] is a general-purpose modelling language, but it includes built-in facilities that allow customizations—or profiles—for a particular domain. A profile fully conforms to the semantics of general UML but specifies additional constraints on selected general concepts to capture domain-specific forms and abstractions. To address this purpose, a formal extension mechanism was defined to allow practitioners to extend the semantics of the UML. The mechanism allows defining stereotypes, tagged values and constraints that can be applied to model elements. A stereotype is an adornment that allows defining a new semantic meaning for a modelled element. Tagged values are key value pairs that can be associated with the modelled element that allow to "tag" any value onto that modelled element. Constraints are rules that define the well-formedness of a model. They can be expressed as text or with the more formal Object Constraint Language (OCL) [20]. Therefore, UML allows with OCL to attach additional information, like invariants, constraints, etc. to the graphical specification [21].

This way, in order to illustrate how a emotional cue can be modelled using defined UML extensions, the user context-aware model of emotion in speech using Ekman's taxonomy for the categorical theory of emotion [5] is described. As it can be seen simplified in Figure 4, the emotion is represented with different speech media properties defined in [22]. Moreover, implied context elements and element properties can be more than the ones shown in Figure 4. Each derived media property is obtained from a basic media property (e.g. fundamental frequency). In the same way, each derived element property is obtained from a basic element property (e.g. gender)

In this application ontology, the information to produce emotional speech is expressed in order to be used for developing speech synthesis systems. Relevant information (i.e. parameters to analyze) to recognize emotions in speech, including context elements to be taken into account, is also shown in this Figure 4. It is noteworthy that this application ontology is useful for developing speech recognition systems (i.e. those presented in [22]) as well.

5 Conclusions and future work

In this paper, an ontology that can be used to describe context-aware emotions is presented. This formalization can be used by developers to construct tools for a generic description of emotions that can be personalized to each user, language and culture.

The main expected benefit of this solution is to establish a formal description of emotions in an easily understandable way, using concepts defined in the ontology. Moreover, proposed ontology enables more abstract description of emotions in various ways. As it can be seen in Figure 1, recognition of emotions can be viewed as a bottom-up process. Firstly, basic media properties are detected, and then derive other media properties. By analyzing derived media properties, emotional cues are

obtained, and by combining these cues, the representation of emotions, taking user context and theory into account. On the other hand, production of emotions can be viewed as a reverse, top-down process.

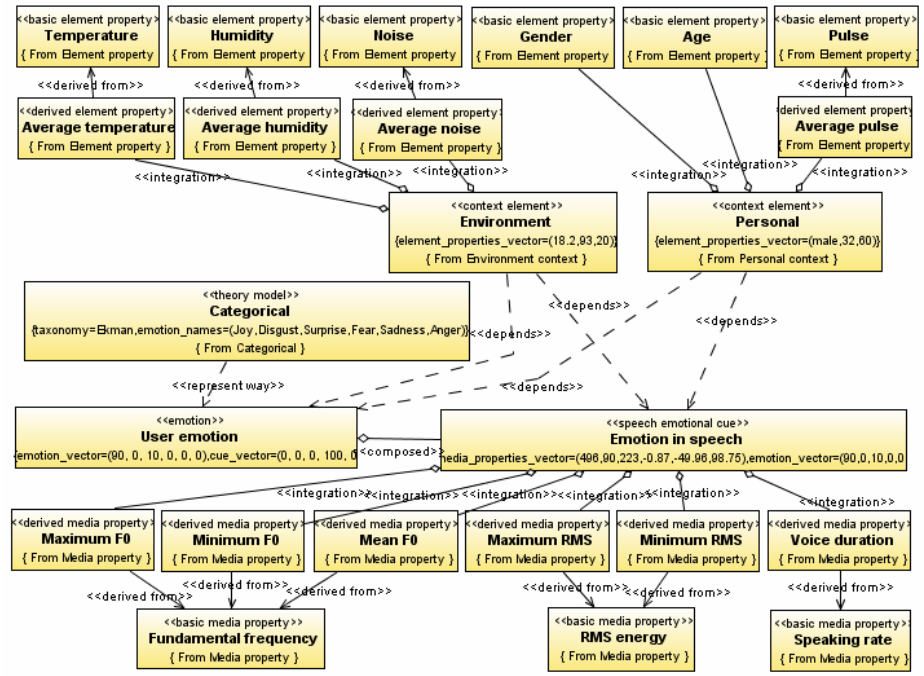

Fig. 4. An example application ontology for describing speech emotional cue, depending on the user context and represented according to Ekman's taxonomy.

This ontology can be used for various purposes. For example, it can serve as a description or metadata about some emotions, or as a part of user profile. Furthermore, most modelling tools allow transformation of models using formats such as Extensible Markup Language (XML) or Web Ontology Language (OWL) so afterwards they can be used by other tools.

Based on the ideas from proposed ontology, authors are working on applications which produce or recognize emotions that do not have "built in" complete knowledge about recognition of contextual emotions. In order to be more flexible they can be parameterized for particular emotional cues, using models of emotions and user context. Generic applications can be created combining emotional cue module and context module, parameterized with developed models. Achieved results are going to be tested in order to establish whether this work will be usable for other people and which level of usability is achieved, in order to maximize it.

So far, authors have mainly described models for the transmission of emotions via speech and some contextual information [3]. Currently, they are also studying emotions according to categorical, dimensional and appraisal models of emotions [4, 5, 6] via other modalities and more contextual information. In the future, more models related to emotion and user context have to be taken into account.

References

1. Picard, R.: Perceptual user interfaces: affective perception, Communications of the ACM, Vol. 43, No. 3 (March 2000) 50-51
2. Norman, D.: Emotional Design. Why we love (or hate) everyday things. Basic Books. (2003). ISBN: 0465051359
3. Obrenovic, Z., Garay, N., López, J. M., Fajardo, I., Cearreta, I.: An ontology for description of emotional cues. In: Tao J., Tan T. and Picard R. W. (Eds.) LNCS 3784 (2005), 505-512
4. Lang, P. J.: A bio-informational theory of emotional imagery. Psychophysiology, Vol. 16, (1979) 495-512
5. Ekman, P.: Expression and nature of emotion. In K. Scherer and P. Ekman (Eds.): Approaches to emotion. Hillsdale, Nueva Jersey: Erlbaum, (1984)
6. Scherer, 1999 K.R. Scherer, Appraisal theory. In: T. Dalgleish and M.J. Power, Editors, Handbook of cognition and emotion, Wiley, New York (1999), pp. 637–663
7. Mehrabian, A., Russell, J. A.: An approach to environmental psychology. Cambridge, MA: MIT Press (1974)
8. Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., Gielen, S.: Acoustic correlates of emotion dimensions in view of speech synthesis. In Proc. Eurospeech, Vol. 1 (2001) 87-90
9. Cowie, R., Douglas-Cowie, E., Cox, C.: Beyond emotion archetypes: Databases for emotion modeling using neural networks. Neural Networks, Vol. 18 (4) (2005) 371-388
10. Dey, A. K., Abowd, G. D.: Towards a better understanding of context and context-awareness. Workshop of The What, Who, Where, When and How of Context-awareness, in CHI2000 Conference, The Hague (The Netherlands) (2000)
11. Göker A., Myrhaug, H. I.: User context and personalisation. ECCBR Workshop on case based reasoning and personalisation. Aberdeen, UK (2002)
12. Model Driven Architecture resource page. Needham, MA: Object Management Group. Retrieved March 08, 2007, from http://www.omg.org/mda/index.htm
13. UMLTM Resource page. http://www.uml.org (2005)
14. Fowler, M, Scott, K.: UML distilled: Applying the standard object modeling language. Addison-Wesley (1997)
15. Kogut, P., Cranefield, S., Hart, L., Dutra, M., Baclawski, K., Kokar, M., Smith, J.: UML for ontology development, The Knowledge Engineering Review, Vol. 17 (1) (2002) 61-64
16. Obrenovic, Z., Starcevic, D.: Modeling Multimodal Human-Computer Interaction, IEEE Computer, Vol. 37 (9) (2004) 65-72
17. W3C Emotion Incubator Group Report 10 July 2007. In Schroder, M., Zovato, E., Pirker, H., Peter, C., Burkhardt, F. (Eds) (2007)
18. McIntyre, G., Göcke, R.: Towards Affective Sensing. In Proc. .HCII 2007, Vol. 3, LNCS_4552, ISBN: 978-3-540-73108-5
19. Obrenovic, Z., Starcevic, D., Devedzic, V.: Using Ontologies in Design of Multimodal User Interfaces. In M. Rauterberg, M. Menozzi, and J. Wesson (Eds.): Human-Computer Interaction - INTERACT '03, IOS Press & IFIP (2003) 535-542
20. Object Constraint Language Specification, version 2.0. Retrieved March 08, 2007, from http://www.omg.org/technology/documents/formal/ocl.htm
21. Oechslein, C., Klügl, F., Herrler, R., Puppe, F.: UML for Behavior-Oriented Multi-Agent Simulations. In: Dunin-Keplicz, B., Nawarecki, E.: From Theory to Practice in Multi-Agent Systems, CEEMAS 2001 Cracow, Poland, September 26-29 (2001) (LNCS 2296), Springer, Heidelberg, pp. 217
22. Álvarez, A., Cearreta, I., López, J.M., Arruti, A., Lazkano, E., Sierra, B., Garay, N.: Feature Subset Selection based on Evolutionary Algorithms for automatic emotion recognition in spoken Spanish and Standard Basque languages. In P. Sojka, I. Kopecek and K. Pala: Text, Speech and Dialog. LNAI 4188, Springer (2006) 565-572