Emotional Face Expression Profiles Supported by Virtual Human Ontology

A. García-Rojas, F. Vexo, D. Thalmann[1], A. Raouzaiou, K.Karpouzis,

S. Kollias[2], L. Moccozet and N. Magnenat-Thalmann[3]

[1]VRLab - EPFL

CH-1015 Lausanne, Switzerland.

email: {alejandra.garciarojas, frederic.vexo, deniel.thalmann}@epfl.ch

[2]Image, Video and Multimedia Laboratory - National Technical University of Ahtens

9, Heroon Politechniou street

15773, Zographou, Athens, Greece

email: {araouz, kkarpou, skollias}@image.ntua.gr

[3]MIRALab - University of Geneva

1211 Geneva, Switzerland

email: {moccozet, thalmann}@miralab.unige.ch

Abstract

Expressive face animation synthesis of human like characters has had many approaches with good results. MPEG-4 standard has functioned as the basis of many of those approaches. In this paper we want to lay down the knowledge of some of those approaches inside an ontology in order to support the modeling of emotional face animation in virtual humans (VH). Inside this ontology we present MPEG-4 facial animation concepts and its relation with emotion through expressive profiles. This relation in the presented profiles is made using psychological models of emotions. The ontology allows storing, indexing and retrieving prerecorded synthetic facial animations that can express a given emotion. Also this ontology can be used as expertise knowledge base of the emotional facial animation creation. This ontology is made using Web Ontology Language and the results are presented as answered queries.

Keywords: ontology, MPEG-4 facial animation, emotion expression

Introduction

The real world actions of a human can be transferred into a virtual environment through a representative (avatar), while the virtual world perceives these actions and responds through respective system avatars which may express their emotions using human-like expressions and gestures. Authors of human face animation synthesis are close attached to MPEG-4 standard because it provides an structured control of the character's face. This paper is based on the work of Raouzaiou et al. who have presented in [1] approaches to synthesize MPEG-4 facial expressions based on discrete and dimensional emotion representation models [2]. The advantage of mentioned approaches described is that real and naturalistic data captured with emotion induction techniques are used and, as a result, the obtained measurements are far more realistic that those portrayed in most facial expression databases, such as MIT's or Ekman's.

The potential and the results of the knowledge posted by those approaches are still in a black box, in the sense that they are described in proprietary rule-based forms, inconsistent with emerging knowledge representation concepts. In this paper we propose to lay down this knowledge for common understanding using ontologies. We aim at structuring concepts involved in the work done for animating facial expressions to provide higher descriptors. Synthesis facial animation comprehends many fields of knowledge: psychology, animation control, standards, etc. and using an ontology suggest a good solution for a formal specification of a shared knowledge [3].

The MPEG-7 standard caters for the interchangeable description of structured information. In MPEG-7, shape descriptors are proposed in order to support the extraction of low level features from 2D and 3D models. In [4], the proposed MPEG-7 3D shape descriptor computes the histogram of the shape index over the whole 3D surface. Such feature extraction can further be incorporated into a 3D model retrieval system. The basic approach [5] usually consists in providing a target 3D shape and in retrieving from the database 3D shapes that show similarities in the features space with the target shape. An alternate application consists in classifying similar 3D shapes into clusters. Although Virtual Human models involve 3D shapes for the body representation, they contain much higher level semantic features. Therefore specific high-level descriptors are required in order to develop accurate models retrieval systems. Semantics aims at easing the construction, functionality and control of VHs. In [6] an ontology-based approach is used for modeling the knowledge involved in the creation of Virtual Humans. They present the complexity this work carry and a general framework is presented.

To explain the objective of this ontology we describe two kinds of scenarios. The fist one is for retrieving animation files that have been annotated with high level descriptors; for example searching for a face animation that express a particular emotion. This is to promote the creation of reusable, scalable and adaptive content in virtual environments. The second scenario is where the ontology can provide a kind of expertise in the knowledge domain. For example to search for the parameters values needed to generate certain kind of expression like: feature points involved, range of their FAPs values, emotion-activation value, etc.

This paper is organized as follows: the next section presents the ontology domain, which is composed by MPEG-4 definitions an the emotion representation; this is followed by the description of the Face Expression Profiles that is based on previews work on facial expression synthesis; after, we present an example of the population of the ontology with the resulting data of the expression analysis; finally the usage scenarios of the ontology and the conclusions are presented.

Ontology Domain

The goal of the ontologies is to lay down the knowledge and express it in a way which is understandable for humans and computers. We first need to define the knowledge of the Face Animation object of MPEG-4 in order to describe the animation structure of VHs. This structure is used to establish a relation between the animation components and emotional parameters. This emotional parameters are modeled using computational models of emotions.

The ontology would help to provide the right animation for a desired expression of emotion. To illustrate this we provide the following questions we aim at answering: What are the face animations for expressing sad emotion? What kinds of anger expression do we have? What is the range of FAP values for the emotion worried, terrified? Given a set of FAP values what is the emotion that can be produced?

In the following subsections explain how is modeled this ontology domain. The ontology is created by using Ontology Web Language [7]. OWL is a formal language that precisely specifies the semantic relationships among entities; it facilitates greater machine interpretability of web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics. This language offers more ways of entity relation, cardinality relation between entities and more others logical operations when compared with older languages. The ontology design was made using Protg ontology editor [8] which is an open source and knowledge-base framework.

MPEG-4 Face Animation Object

MPEG-4 has been exploited because it is possible to have a structured control of the character. Some of the standard advantages are: very low bit rate coding, high quality and customizable rendering; scalability, etc.

In the framework of MPEG-4 standard, some parameters have been specified for Face and Body Animation (FBA). For face definition MPEG-4 specifies 84 feature points on the neutral face [9], which provide spatial reference for the Facial Animation Parameters (FAPs) definition. The FAP set contains two high-level parameters, visemes and expressions. Facial shape is defined by the Facial Definition Parameters (FDP) and Facial Deformation Tables (FDT) nodes. FDP contains the feature points in the face shape and the FDT contains the deformation of the model face as a function on its animation. For animating a face FAP node has the translations of feature points, expressed in FAP Units; they can produce high quality facial expressions and speech pronunciation. FAPs that can be used together are represented in groups to facilitate the animation [10]. This structure of nodes is represented in the ontology,shown in Figure 1.

Emotional Expression Representation

Depending on the context of interaction, one prefer to choose a discrete emotion representation models, such as Ekman's, over a dimensional. For example, if the objective of an application is classification of images or videos into specific categories, it is better to use discrete representations. On the other hand, dimensional models perform better when capturing subtle emotions (everyday human discourse), and can be differentiated between the two principal axes, e.g. in an anger detection application used in automated phone centers. Besides these representations, component process models, such as Stacy-Marsella's or Scherer's appraisal checks, aim at mimicking human brains processes when confronted with new events. In these situations, several aspects of a stimulus are taken into account, such as its disposition against the aims of the subject, its novelty, etc. Albeit they are useful in modeling human-human discourse, there are several unclear aspects in these models, such as timing or sequencing of processes, which hamper the development of related computational models.

VH's face is capable of having explicit synthetic expressions by itself within animations [11] [12] according to a given emotion. Ekman's model of emotion [13] has been widely accepted to describe the six archetypal emotions. Within MPEG-4 for each of these basic emotions a list of FAPs is associated (those FAPs are animated to produce the specified emotion). But this model is not sufficient to produce a varied range of emotions; therefore we need to incorporate a more transparent and continuous representation of emotions. Whissel's activation-evaluation space [14] is a simple circular representation that captures a wide range of emotions and simplifies them in two dimensions: activation and evaluation. As shown in Figure 2 the vertical axe represents the activation value and the horizontal one the evaluation value.

The ontology diagram in the figure 3 shows that one emotion can be modeled by one or more Models of Emotions. Universal emotions, as well as intermediate ones, are represented as points in the activation-evaluation space as is shown in figure 2.

In next section we will describe prior work on facial expression synthesis that group facial expression inside profiles. This profiles are close related with the mentioned models of emotion.

Face Expression Profiles

As general rule, one can define six general categories for facial expressions, each categorized by an archetypal emotion. Within these categories, expressions are described by different emotional intensities, as well as minor variations in expression details. The need for different Expression Profiles [1] arises from the fact that the same emotion may be expressed via the face using different facial deformations, subject to personal traits or different flavors in expressivity. This concept is useful in both analysis and synthesis purposes, since it allows for personalized, diverse expressions, thus reducing the possibility of robot-like faces. In the following subsections we will explain the Profiles defined for archetypal and intermediate facial expressions. These profiles compound the knowledge we aim at representing in the ontology.

Archetypal Expression Profile

FAPs representations make good candidates for describing quantitative facial motion features. The use of these parameters serves several purposes such as compatibility of created synthetic sequences with the MPEG-4 standard and increase of the range of the described emotions. In general, archetypal expressions can be uniformly recognized across cultures and they are therefore invaluable in trying to analyze the user's emotional state. These expressions occur rather infrequently and in most cases emotions are expressed through variation of a few discrete facial features related with particular FAPs.

Based on elements from psychological studies and from statistical analysis [1], we have described the six archetypal expressions using MPEG-4 FAPs, and 25 archetypal profiles were identified. An example of archetypal expressions sadness and fear is illustrated in Table 1 where we can see the FAPs involved to produce these expressions. Table 2 presents the profiles created for these expressions considering ranges of FAP values for the same expression. In this case Sadness has only one profile and Fear has nine. The complete list of all archetypal expression profiles and their FAPs ranges can be found in [1]. In the case of an emotion that has many profiles it means that this emotion can be expressed in different ways. For example the fear emotion profile zero has this characteristics.... (+++++ explain a difference between profiles ++++++) and the fear emotion profile four has this other characteristics (+++++).

Intermediate Expression Profile

Creating profiles for an expression that cannot be clearly characterized as an archetypal one is not straightforward. Apart from estimating the range of variations for FAPs, one should first define which FAPs are involved in the particular expression.

One is able to synthesize intermediate expressions by combining the FAPs employed for the representation of two archetypal ones. In our approach, in order to define the profiles of intermediate expressions we use Whissel's wheel [14] which suggests that expressions are points in a space. FAPs that are common in both expressions are retained during synthesis, while FAPs used in only one expression are averaged with the respective neutral position. In the case of having FAPs in common, averaging of intensities usually favors the most exaggerated of the expressions that are combined, whereas FAPs with contradicting intensities are canceled out.

This is a rule-based technique for analysis and synthesis to merge profiles of archetypal expressions for creating profiles of intermediate. An example of the Intermediate expression depressed is presented in the Table 3 the range variation resulting of combining Sad ($P_S^{(0)}$) and Fear ($P_F^{(4)}$) archetypal profiles using the activation-evaluation measures. Not all the FAPs are presented in this table, but for a detailed explanation and reference to intermediate expressions look at [1]. An example of how two archetypal profiles can be merged, if we consider the sad profile that has this characteristics (+++ explain +++)which is combined with the profile Fear one that has (+++ explain +++) the result is an expression with the characteristics (+++ explain +++).

Ontology Of Facial Expression Animation

As already described, we have defined two types of Profiles for facial expression: Archetypal and Intermediate. In Figure 4 we present that a facial expression can be archetypal or intermediate; and all facial expression profiles have a range of FAP values according to the rules defined above. The intermediate profile encloses two archetypal expressions and provides a FAPs Range Variation. For intermediate expression we made restrictions in the ontology based on the rules we followed to create them. To describe this knowledge we present the next assertions:

1. The Face expression can be defined by a profile Archetypal or Intermediate.

2. Each Intermediate Expression should enclose 2 archetypal expressions.

$IntermediateProfile\ enclosesArchetypalProfiles\ =\ 2$

3. Each profile has a Range of FAPs associated (FAPRange). The FAPsRange class contains the max and min FAPs values.

4. The Range of FAPs of Intermediate expressions is the union of the Ranges of the Archetypal expression that are enclosed by the Intermediate expression. This can be expressed as a restriction for the hasFAPsRange property as:

$\forall\, hasFAPsRange(FAPRange \bigcap ((\exists\, isRangeOfExpressionProfile\ ArchetypalProfile) \bigcap (isEnclosedBy\ IntermediateProfile)))$

5. Intermediate expressions FAP Range Variation is computed considering three possibilities. a) Mutually inclusive FAPs with different sign b) Mutually inclusive FAPs with same sign c) Mutually exclusive FAPs The computation of these values is through the expression generator (this issue will be explained in the next section in Rule Population, we will also explain how the ontology could be populated and used).

Ontology Rule Population

In this section we give an example of how to populate the expression profiles in the ontology. In Figure 5 we present a diagram of some instances of Fear and Sad archetypal emotional expression and Depression as intermediate expression accompanied by graphical results obtained using the presented method for synthesizing face expression.

Fear emotion has a face expression FaceExpressionFear. This face expression is defined as archetypal, and as a consequence has Archetypal Profiles defined: ArchetypalProfiel_Fear0, ArchetypalProfiel_Fear4. Each profile has their FAP Range values obtained form [1]. Sad archetypal expression (ArchetypalProfiel_Sad0) is defined in the same way.

To define the Depression as intermediate expression, in between of Fear and Sadness, we created the Depression Face Expression as an intermediate expression with its Profile which encloses the Archetypal Expressions Fear and Sadness. The FAPRange of the Depression profile is implicitly declared as the union of the FAP range of Fear and Sad profiles. For this intermediate profile the FAPRangeVariation contains the FAP range values that were obtained from the intermediate expressions generator presented in [15]. In this example we just show the FAP 19 close_t_l_eyelid which is contained in both archetypal expressions and has the same sign, which is the first case of the rules specified in the previous section.

In the same way all the possible facial expressions with their profiles should be integrated in the ontology. In the following section we will present examples of the proposed scenarios in which the ontology aims to be useful.

Ontology Usage Scenarios

The way to extract information from ontologies is by making queries. To do this we have used the plug-in new Racer Query Language interface [16] for OWL ontologies in Protg. This query language is close to natural language. nRQL language is provided by RacerPro [17] which is a system for managing semantic web ontologies based on OWL. RacerPro is necessary to be run for providing reasoning when querying the ontology using nRQL.

Now we present in RQL the questions that were formulated in the Ontology domain section. The first question belongs to the scenario where we can obtain animations that have been annotated with the structure proposed in the ontology. We present the responses to these questions we are able to obtain by populating the ontology.

In the first scenario we want to find animations that express a specific emotion.

What is the facial animation for expressing depressed emotion?

nRQL:

$(retrieve\ (?a)\ (and\ (?a\ |FaceAnimation|)$

$(?a\ ?b\ |hasExpresion|)\ (?b\ |FaceExpression|)$

$(?b\ ?c\ |describesEmotion|)\ (c?\ |Depressed|)))$

Result: $(((?A\ |\ FaceAnimation\_Depressed|))$

We have found that there is the FaceAnimationDepressed defined inside the ontology. In the case of the second scenario probably we want to know the information used for defining this face expression, like: is the expression is intermediate or archetypal?; Which archetypal expressions it encloses? What is its FAP range? Etc.

What is the range of FAP values for the face expression depressed?

nRQL:

$(retrieve\ (?H)\ (and\ (?A\ |FaceAnimation|)$

$(or\ (and\ (?A\ ?B\ |hasExpresion|)\ (?B\ |FaceExpression|)$

$(?B\ (some\ |describesEmotion||CONCEPT-FOR-Depressed|))$

$(?B\ ?F\ |hasProfile|)\ (?F\ |IntermediateProfile|)$

$(?F\ ?G\ |enclosesArchetypalProfiles|)\ (?G\ |ArchetypalProfile|)$

$(?G\ ?H\ |hasFAPsRange|)\ (?H\ |FAPRange|)))))$

Result: $(((?H\ |Sad0_F19|))\ ((?H\ |Fear4_F3|)))$ ...

The last query brings the union of the set of FAPs of the archetypal expressions that are used for the intermediate depressed profile. In Figure 6 we show a screen shot of the nRQL plug-in inside Protégé when submitting a query.

Conclusions and Future work

In this paper we have presented an ontology that lays down the knowledge of previous work on facial animation expression within MPEG-4 framework. The facial expressions have been represented in a form of Archetypal and Intermediate Profiles that describe the synthesis of the emotion. The structure of this ontology allows us to retrieve animations that have been annotated under this structure; and also to retrieve the expertise covered in the synthesis of those facial expressions. The main issue that is normally faced when adding semantics to a knowledge domain is the extraction of the meta-data and the interconnection with existing knowledge sources, such as audiovisual databases with expressive material. This issue needs to be solved as part of our future work.

Another issue that is currently being investigated in the integration of the ontology with an MPEG-4 compliant animation systems such as Greta [11]. In this scenario, the animation would query the ontology for the representation of an avatar with specific expressivity, e.g. a woman showing extreme joy. The result of this query would be the definition of the avatar geometry and a set of facial expressions and body gestures, ready to use in an animation. Part of this work, is the extension of these concepts in gesturing; neutral gestures used in interaction can be transformed into expressive ones by taking into account measurable transformations, such as Laban parameters, while deictic gestures such as shrugging can vividly illustrate emotions without any further processing. Moreover the technology available for processing OWL files is more focused for web as the semantic web; an on-line framework for real time animation is still matter of research. Therefore what we can offer under this conditions is a specific search scenarios for facial expressive animations.

Acknowledgements

This research has been funded by the EC and the Swiss Federal Office for Education and Science in the framework of the European Networks of Excellence IST-HUMAINE (http://emotion-research.net) and IST-AIM@SHAPE (http://www.aim-atshape.net).

References

[1] A. Raouzaiou, N. Tsapatsoulis, K. Karpouzis, and S.D. Kollias. Parameterized facial expression synthesis based on mpeg-4. 2002(10):1021–1038, October 2002.

[2] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor. Emotion recognition in human-computer interaction. In *IEEE Signal Processing Magazine, no. 1, January 2001, pp.32-80*, 2001.

[3] T. R. Gruber. The role of a common ontology in achieving sharable, reusable knowledge bases. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, pages 601–602, 1991.

[4] Titus, Zaharia, Franois Prteux, and Marius Preda. 3d shape spectrum descriptor.

[5] T. Zaharia and F. Prteux. Three-dimensional shape-based retrieval within the mpeg-7 framework. In *SPIE Conference 4304 on Nonlinear Image Processing and Pattern Analysis XII*, pages 133–145, San Jose, CA, January 2001.

[6] Mario Gutiérrez, Daniel Thalmann, Frédéric Vexo, Laurent Moccozet, Nadia Magnenat-Thalmann, Michela Mortara, and Michela Spagnuolo. An ontology of virtual humans: incorporating semantics into human shapes. In *Proceedings of Workshop towards Semantic Virtual Environments (SVE05), March 2005*, pages 57–67, 2005.

[7] W3C. Owl web ontology language. http://www.w3.org/tr/owl-features/.

[8] Protégé. (c) 2005 stanford medical informatics. http://protege.stanford.edu/index.html.

[9] A.M. Tekalp and J. Ostermann. Face and 2-d mesh animation in mpeg-4. 15(4-5):387–421, January 2000.

[10] ISO/IEC JTC1/SC29/WG11 N4668. Overview of the mpeg-4 standard, March 2002.

[11] Fiorella de Rosis, Catherine Pelachaud, Isabella Poggi, Valeria Carofiglio, and Berardina De Carolis. From greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *Int. J. Hum.-Comput. Stud.*, 59(1-2):81–118, 2003.

[12] Koray Balci. Xface: Open source toolkit for creating 3d faces of an embodied conversational agent. In *Smart Graphics*, pages 263–266, 2005.

[13] Eckman P. Facial expression ans emotion. *Am Psychologist*, 48:384–392, 1993.

[14] C. M. Whissel. *The dictionary of affect in language R. Plutchnik and H. Kellerman (Eds) Emotion: Theory, research and experience: vol 4, The measurement of emotions*. Academic Press, New York, 1989.

[15] A. Raouzaiou, E. Spyrou, K. Karpouzis, and S. Kollias. Emotion synthesis: an intermediate expressions generator system in the mpeg-4 framework. International Workshop VLBV05, 15-16 September 2005, Sardinia, Italy, 2005.

[16] nRQL. new racer query language (nrql) interface to owl ontologies in protégé. http://www.cs.concordia.ca/k̃_bhoopa/nrql.html.

[17] RacerPro. Renamed abox and concept expression reasoner. http://www.racer-systems.com.

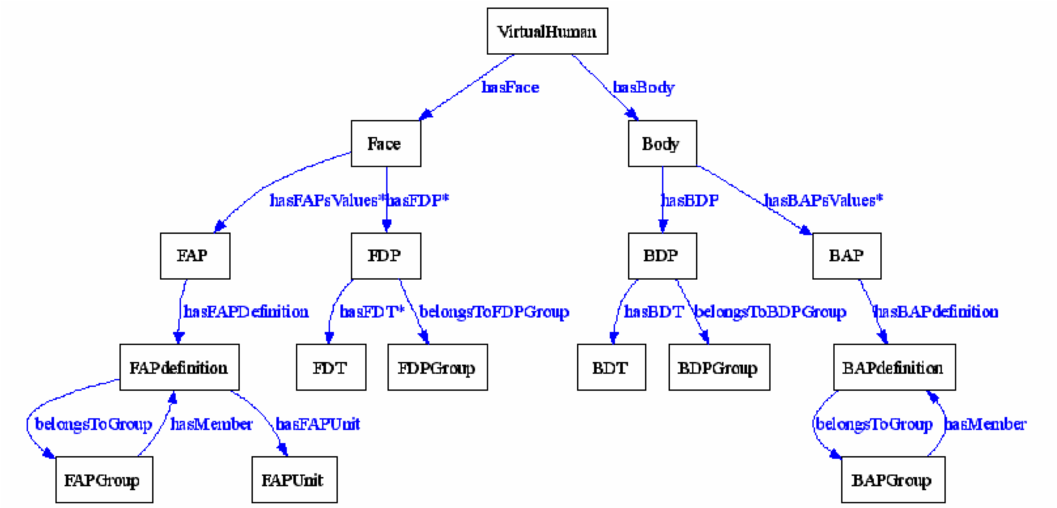

Figure 1: Representation of the Facial and Body Animation object of MPEG-4 in the ontology.

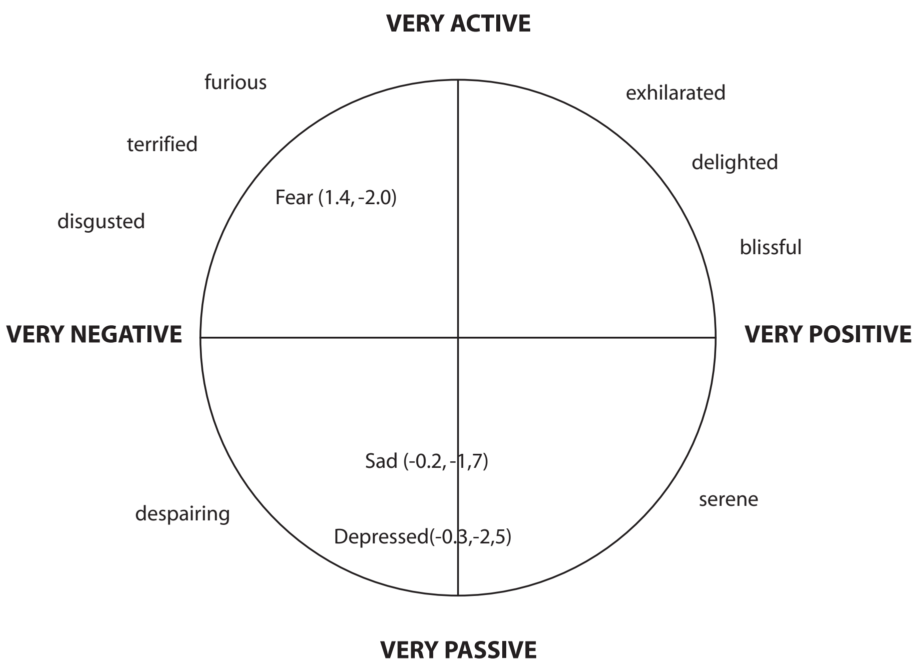

Figure 2: Activation-Evaluation space.

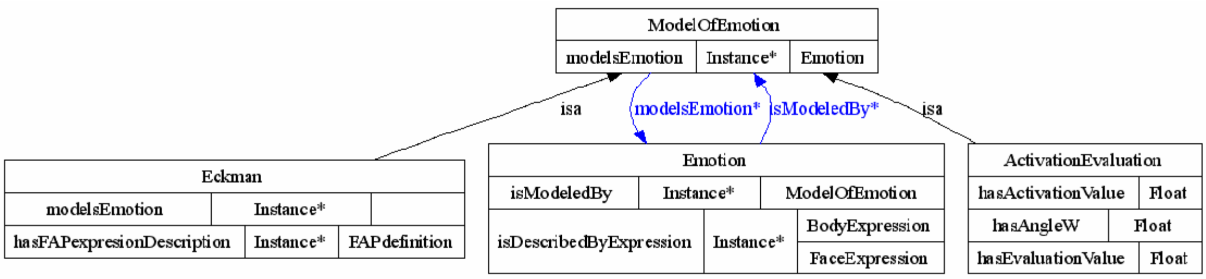

Figure 3: Description of the psychological models of emotion in the ontology.

Sadness	close_t_l_eyelid(F_{19}), close_t_r_eyelid(F_{20}), close_b_l_eyelid(F_{21}), close_b_r_eyelid(F_{22}), raise_l_i_eyebrow(F_{31}), raise_r_i_eyebrow(F_{32}), raise_l_m_eyebrow(F_{33}), raise_r_m_eyebrow(F_{34}), raise_l_o_eyebrow (F_{35}), raise_r_o_eyebrow(F_{36})
Fear	open_jaw(F_3), lower_t_midlip(F_4), raise_b_midlip(F_5), lower_t_lip_lm(F_8), lower_t_lip_rm(F_9), raise_b_lip_lm(F_{10}), raise_b_lip_rm(F_{11}), close_t_l_eyelid (F_{19}), close_t_r_eyelid(F_{20}), close_b_l_eyelid (F_{21}), close_b_r_eyelid(F_{22}), raise_l_i_eyebrow(F_{31}), raise_r_i_eyebrow(F_{32}), raise_l_m_eyebrow(F_{33}), raise_r_m_eyebrow(F_{34}), raise_l_o_eyebrow(F_{35}), raise_r_o_eyebrow(F_{36}), squeeze_l_eyebrow(F_{37}), squeeze_r_eyebrow(F_{38})
...	...

Table 1: FAPs Vocabulary for Archetypal Expression Sadness and Fear.

Profiles	FAPs and Range of Variation
Sadness (P_S^0)	F_{19} ∈[-265,-41], F_{20} ∈[-270,-52], F_{21} ∈[-265,-41], F_{22} ∈[-270,-52], F_{31} ∈[30,140], F_{32} ∈[26,134]
Fear ($P_F^{(0)}$)	F_3 ∈[102,480], F_5 ∈[83,353], F_{19} ∈[118,370], F_{20} ∈[121,377], F_{21} ∈[118,370], F_{22} ∈[121,377], F_{31} ∈[35,173], F_{32} ∈[39,183], F_{33} ∈[14,130], F_{34} ∈[15,135]
...	...
Fear ($P_F^{(4)}$)	F_3 ∈[400,560], F_5 ∈[-240,-160], F_{19} ∈[-630,-570], F_{20} ∈[-630,-570], F_{21} ∈[-630,-570], F_{22} ∈[-630,-570], F_{31} ∈[460,540], F_{32} ∈[460,540], F_{33} ∈[360,440], F_{34} ∈[360,440], F_{35} ∈[260,340], F_{36} ∈[260,340], F_{37} ∈[60,140], F_{38} ∈[60,140]
...	...

Table 2: Profile for the Archetypal Expression Sadness and Fear.

Expression	**Activ-Eval**	F_3**min**	F_3**max**	F_5**min**	F_5**max**	F_{19}**min**	F_{19}**max**	...
Fear ($P_F^{(4)}$)	$(1.4, -2.0)$	400	560	-200	-160	-630	-570	...
Depressed ($P_D^{(0)}$)	$(-0.3, -2.5)$	160	230	-100	-65	-110	-310	...
Sad ($P_S^{(0)}$)	$(-2.0, 1.7)$	0	0	0	0	-265	-41	...

Table 3: Intermediate Profile of Depression and the archetypal expressions used to calculate it using the activation and evaluation measures.

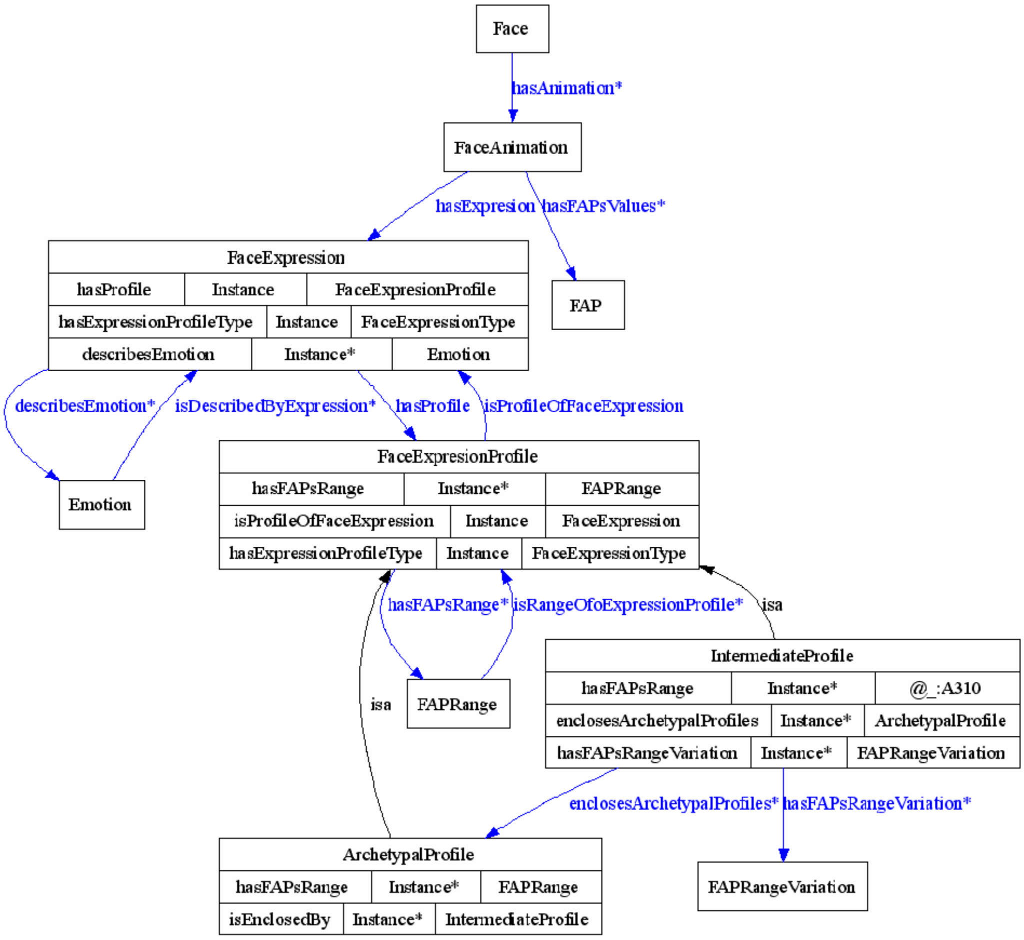

Figure 4: Description of the archetypal and intermediate face expression profiles.

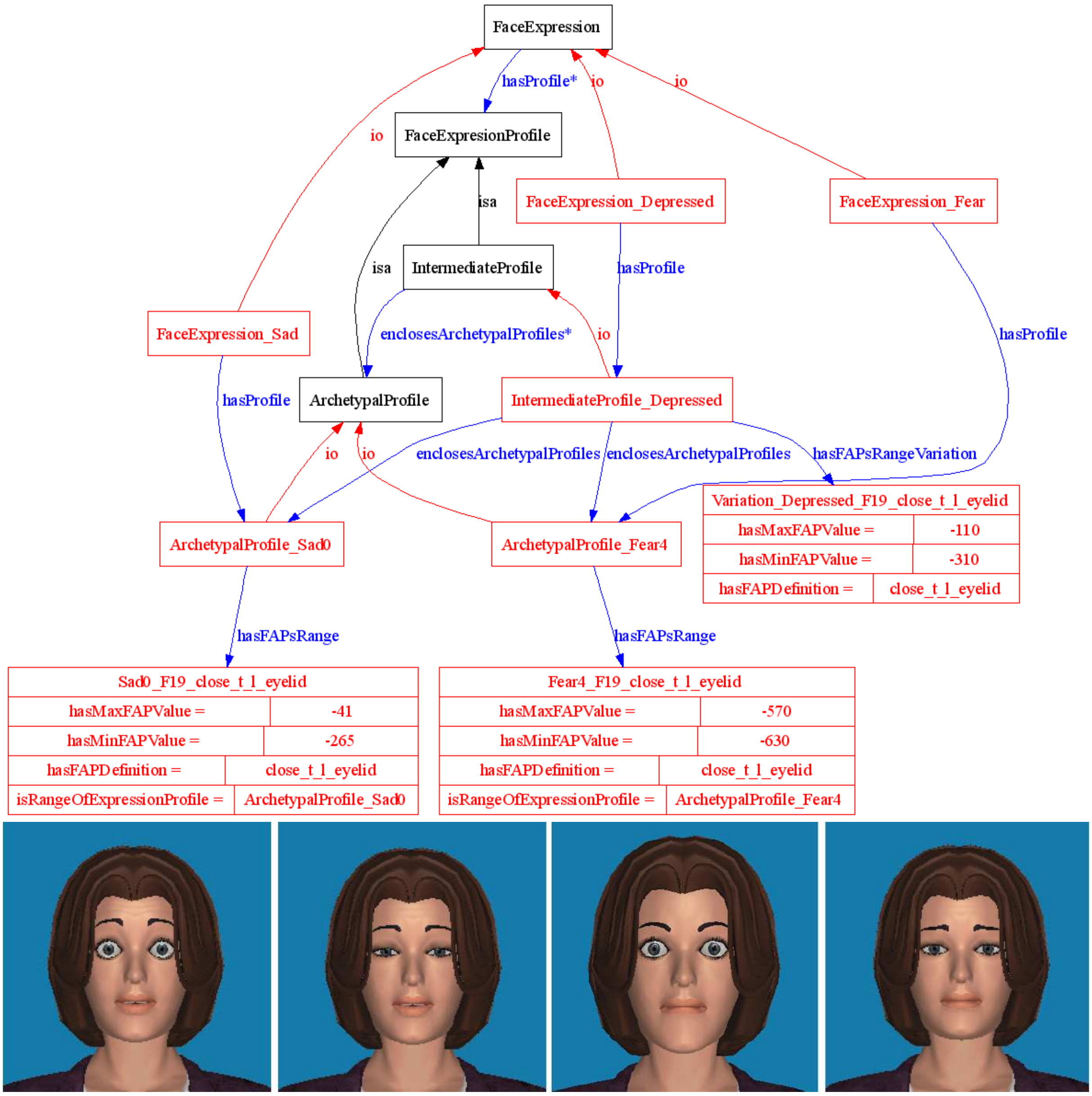

Figure 5: Example of face expression profile population and the graphical results for 1-fear, 2 and 3 - depress and 4 - sadness expressions.

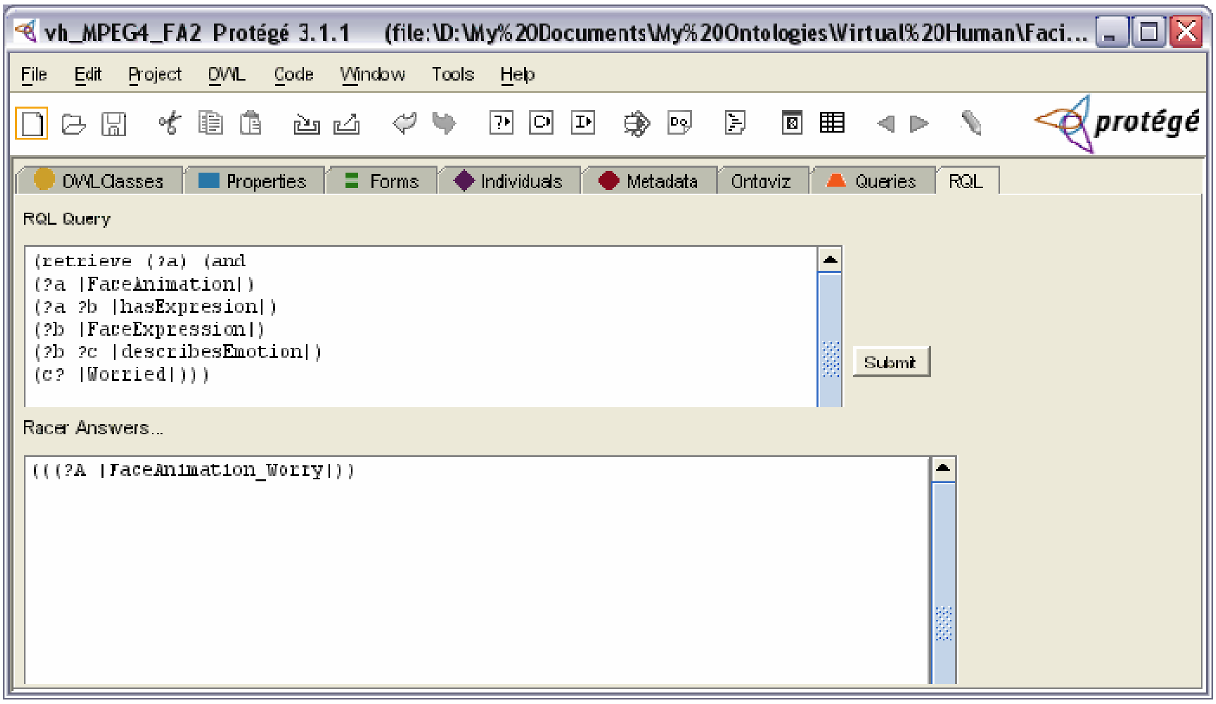

Figure 6: Snapshot of Protégé using nRQL plug-in to make queries.