KRISTINA: A Knowledge-Based Virtual Conversation Agent

Leo Wanner[1,2], Elisabeth André[3], Josep Blat[2], Stamatia Dasiopoulou[2], Mireia Farrùs[2], Thiago Fraga[4], Eleni Kamateri[5], Florian Lingenfelser[3], Gerard Llorach[2], Oriol Martínez[2], Georgios Meditskos[5], Simon Mille[2], Wolfgang Minker[6], Louisa Pragst[6], Dominik Schiller[3], Andries Stam[7], Ludo Stellingwerff[7], Federico Sukno[2], Bianca Vieru[4], and Stefanos Vrochidis[5]

[1]ICREA, [2]Universitat Pompeu Fabra, [3]Universität Augsburg, [4]Vocapia Research, [5]CERTH, [6]Universität Ulm, [7]Almende

Abstract. We present an intelligent embodied conversation agent with linguistic, social and emotional competence. Unlike the vast majority of the state-of-the-art conversation agents, the proposed agent is constructed around an ontology-based knowledge model that allows for flexible reasoning-driven dialogue planning, instead of using predefined dialogue scripts. It is further complemented by multimodal communication analysis and generation modules and a search engine for the retrieval of multimedia background content from the web needed for conducting a conversation on a given topic. The evaluation of the 1st prototype of the agent shows a high degree of acceptance of the agent by the users with respect to its trustworthiness, naturalness, etc. The individual technologies are being further improved in the 2nd prototype.
Keywords. conversation agent, multimodal interaction, ontologies, dialogue management.

1 Introduction

The need for intelligent conversation agents as social companions that are able to entertain, coach, converse, etc. with those who feel, e.g., lonely or overstrained is on the rise. However, in order to be able to act as a social companion, an agent must be eloquent, knowledleable, and possess a certain cultural, social and emotional competence. Considerable advances have been made to increase the agent's affective and social competence; see., e.g., [37, 26, 4]. However, most of the current proposals in the field still do not rise up to the challenge as a whole. Thus, they usually follow a predefined dialogue strategy (which cannot be assumed when interacting with, e.g., elderly); they do not take into account cultural idiosyncrasies of the addressee when planning their actions; they are not multilingual to be able to intermediate between a migrant and a native from the host country; etc. See, e.g., [1, 36] for some representative examples. To essentially improve on the capacity of a conversational agent to conduct a versatile emotionally and culturally sensitive dialogue, the role of the knowledge model underlying the agent must be reconsidered. An advanced ontology-based

knowledge model is capable of capturing the content of the multimodal (verbal, facial, and gestural) communication input of the user in terms of abstract interpretable structures. Furthermore, it facilitates the interpretation of the input of the user and the decision on the next move of the agent by means of a variety of reasoning mechanisms. And, obviously, it also facilitates the dialogue history bookkeeping, the representation of the cultural and social specifics of a user, as well as domain-specific and common sense knowledge.

In what follows, we present the design and first prototypical implementation of an agent (henceforth referred to as "KRISTINA") in which the knowledge model is central. KRISTINA is projected as an embodied companion for (elderly) migrants with language and cultural barriers in the host country and as a trusted information provision party and mediator in questions related to basic care and healthcare. Consider an excerpt of a sample dialogue as targeted by KRISTINA:

K: *You look downhearted today.What is wrong?*
U: *I feel sad. Because of my eyes, I even can't read the newspaper anymore.*
K: *Shall I read the newspaper aloud for you?*
U: *Yes, this would be great!*
K: *You certainly can still read the headings of the articles. Just tell me which one I shall read.*
... ...

2 Architecture of the KRISTINA agent

Figure 1 shows the global design of KRISTINA, which is targeted to have the following characteristics embedded into linguistic, cultural, social and emotional contexts: **(i)** to be able to retrieve multimedia background content from the web in order to show itself informed and knowledgeable about themes relevant to the user; **(ii)** understand and interpret the concerns of the user expressed by a combination of facial, gestural and multilingual verbal signals; **(iii)** plan the dialogue using ontology-based reasoning techniques in order to be flexible enough and react appropriately to unexpected turns of the user; **(iv)** communicate with the user using verbal and non-verbal (facial and gestural) signals.[1]

The agent is composed of a collection of modules that ensure informed multimodal expressive conversation with a human user. The communication analysis modules are controlled by the *Social Signal Interpretation* (SSI) framework [34]. SSI supports audio and video signal streaming and realtime recognition, synchronization, analysis and high level fusion of the different modality signals within these streams: emotional speech, mimics, and head and body gestures. In KRISTINA, SSI is the central instance for the analysis and synchronization of video and audio signals with respect to displayed emotions. For this purpose, targeted machine learning modules for paralinguistic, facial and gesture analysis

[1] Due to the lack of space, we cannot present a complete run of an interaction turn. Therefore, we merely introduce in what follows the individual modules and sketch how they interact.

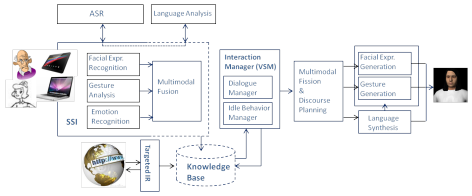

Fig. 1. Architecture of the KRISTINA agent

have been implemented as SSI components–which also ensures seamless interaction with the rest of the framework. For the linguistic analysis of the audio, the transcribed material is piped through SSI to the language analysis module.

The semantic structures obtained from the analysis modules are handed over by the dialogue manager (DM) to the knowledge integration (KI) module in order to be projected onto genuine ontological (OWL) structures, fused and stored in the knowledge base (KB). The dialogue-oriented modules are embedded in the *Visual Scene Maker* (VSM) framework [12]. While the original purpose of VSM has been to support the definition of the interactive behavior of virtual characters, we use it, on the one hand, as a communication shell between the DM module and the modules it interacts with, and, on the other hand, for modeling the idle behavior of the agent.

The DM chooses the best system reaction (in terms of ontological structures), in accordance with the analyzed user move, the user's emotion and culture and the recent dialogue history. For this purpose, it solicits first from the KI module possible reactions that are reasoned over the KB. In other words, in contrast to most of the state-of-the-art DM models, the determination of the turn of the system is distributed between a high level control DM and a reasoning KB module.

The ontological structures of the best system reaction are passed by the DM to the fission (or modality selection) and discourse planning module, which shall ensure an adequate assignment of the content elements chosen for communication to the individual modalities (voice, face, and body gesture) and their coherent and coordinated presentation. The three modality generation modules determine the form of their respective content elements. The language generation module feeds its intermediate and final outcome also to the facial expression and gesture generation modules in order to ensure, e.g., accurate lip synchronization and beat gestures of the virtual character.

A dedicated search engine acquires background multimodal information from the web and relevant curated information sources. The engine extracts content from web resources (including social media) to enhance the background knowledge of KRISTINA that is stored in the KB in terms of ontologies, which facilitates the realization of flexible reasoning-based dialogue strategies.

3 The Knowledge Model of the KRISTINA Agent

To ensure that the agent is "knowledgeable" about the topic of the conversation and thus able to interpret the multimodal input (question, comment, request, etc.) of the user and come up with the appropriate reaction, the knowledge representation in the agent must be theoretically sound and scalable. The knowledge repositories must separate the representation of the state of an ongoing conversation from the high level typology of the conversation (or dialogue) acts and be dynamically extendable, i.e., the agent must be able "to learn" from both the input of the user and the external world.

3.1 Knowledge Representation, Integration and Interpretation

KRISTINA's multimodal knowledge representation framework includes ontologies designed to support the dialogue with the user and to represent the relevant basic care and healthcare background information from the web. The ontologies cover: (i) models for the representation, integration and interpretation of verbal and non-verbal aspects of user communication piped in by the DM [27]; (ii) domain models that capture the various types of background knowledge, including user profile ontologies [15]; ontologies for modeling routines, habits and behavioural aspects [25], and healthcare and medical ontologies [28].

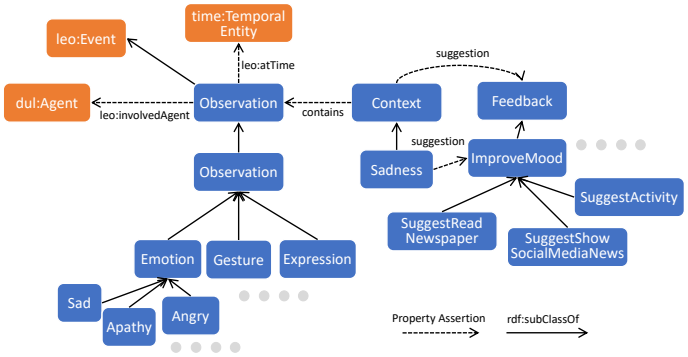

Fig. 2. Observation and Context models

The knowledge integration and interpretation models define how the structures can be combined to derive high-level interpretations. To achieve this, a lightweight ontology pattern is provided for capturing contextual semantics, i.e., the types of the structures that are of interest and the way they should be interpreted by the ontology reasoning task. Figure 2 depicts the vocabulary used for the interpretation of the user's statement *I feel sad* and the complementary information from the visual channel 'low mood' detected via the corresponding valence/arousal values [27]. The ontology extends the `leo:Event` concept of

LODE [32] to benefit from existing vocabularies for the description of events and observations. Property assertions about the temporal extension of the observations and the agent (actor) are allowed, reusing core properties of LODE. The figure also depicts the relationship between observation types and context models in terms of the `Context` class, which allows one or more `contains` property assertions referring to observations.

In our example, the fact that the user is sad constitutes contextual information that is modeled as an instance of `Context`, which is further associated with an instance of `Sad`.

```
:sad1 a :Sad ;
  leo:atTime :t1 ;
  leo:involvedAgent [a dul:Agent].
:t1 a time:TemporalEntity ;
  time:hasBeginning [a time:Instant ;
    time:inXSDDateTime "2017-01-02T18:06:46"];
  time:hasEnd [a time:Instant ;
    time:inXSDDateTime "2017-01-02T18:06:51"].
:ctx1 a :Context;
  :contains :sad1 .
```

Figure 2 also displays an excerpt of the domain ontology used to infer feedback and suggestions based on the emotional state of the user. For each context, one or more `suggestion` property assertions can be defined to associate it with feedback instances that can improve user's mood. In our example, `Sadness` is a subclass of `Context`, defined in terms of the following equivalence axiom:

$$Sadness \equiv Context \sqcap \exists contains.Sad$$

It also defines a property restriction that specifies the type of feedback needed when this emotional context is detected:

$$Sadness \sqsubseteq \exists suggestion.ImproveMood$$

As such, the `ctx1` instance of the example is classified in the `Sadness` context class, which further inherits the restriction about the potential feedback that could be given to improve the mood of the user. All three subclasses of the `ImproveMood` concept are retrieved and sent back to the DM in order to finally select the one that should be returned to the user.

3.2 Dialogue act representation

As already mentioned in Section 2, the DM is responsible for choosing the best suited system action among the suggestions of the KI module. Different aspects, such as the user's emotion and culture as well as the recent dialogue history, are taken into account. The rule-based choice is grounded in the dedicated model of dialogue acts shown in Figure 3.

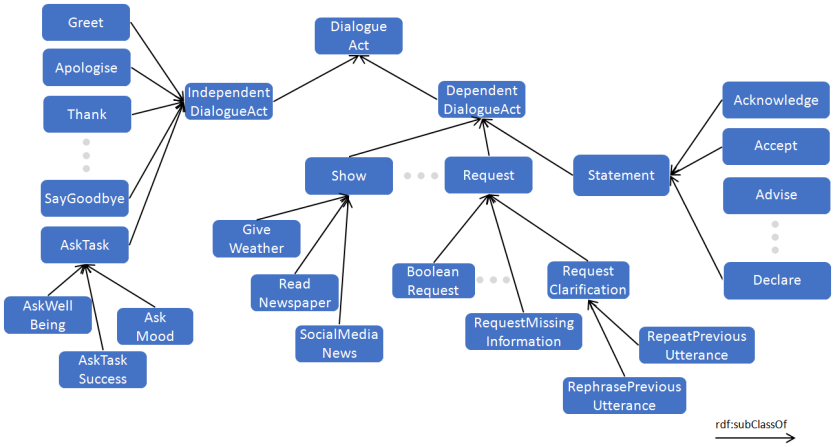

Fig. 3. Excerpt of the dialogue acts ontology

In order to avoid the predefinition of all user and system actions and be able to handle arbitrary input from both the language analysis and the KI modules, the rules are not defined for specific actions, but rather for general features such as the respective dialogue act and the topics, constituted by the classes associated with the possible system actions. For instance, in our example, three system actions are available. They share the dialogue act `Statement`. However, the topics differ. Thus, the first action has the topics `newspaper` and `read`, the second `socialmedia` and `read`, and the third `activity`. Individuals from a collectivistic culture tend to be more tightly integrated in their respective social groups, while individuals from an individualistic culture less so [16]. Therefore, the DM would propose to the user with a collectivistic culture background to read aloud news from social media, and select one of the other options if the user's culture is individualistic.

4 Multimodal interaction

With the knowledge model as its core, the KRISTINA agent performs the entire multimodal interaction, which involves dialogue management and multimodal communication analysis and generation.

4.1 Dialogue management

Besides the maintainance of the dialogue state and selection of the next system action sketchd in the previous section, dialogue management deals with the control of the agent's turn-taking behavior and the control of a variety of non-verbal idle behavior patterns [22]. To manage these two tasks, we use the *Visual SceneMaker* (VSM) platform [12, 21]; see also Section 2. VSM determines the agent's

participant role changes during the dialogue, based on the observed user input and the agent's own actions selected by the DM. The turn-taking decisions are made on the basis of a policy that determines whether the agent is allowed to interrupt the user's utterance and how it reacts to the user's attempts to barge in in its own turn. VSM is also responsible for planning appropriate and vivid non-verbal behavior patterns while the agent is listening to the user or whenever the speaker and listener roles are not yet clearly negotiated. In this latter case, the agent fulfills the role of a bystander by displaying an idle behavior that is supposed to create an impression of engagement and attentiveness while waiting for the user's next dialogue move or before actively starting a contribution itself, for example, mimicking the user's affective state by mirroring their facial expressions, gestures or body postures or displaying different eye gazes [23].

4.2 Multimodal communication analysis

The objective of multimodal communication analysis is to convert the verbal and affective information captured from the user into abstract representations that are projected onto ontologies.

The analysis of verbal (spoken) communication consists of two major tasks: speech recognition and language analysis.[2] For speech recognition, we use the Vocapia ASR[3], which exploits statistical speech models for both acoustic and language modeling [19]. Language analysis captures the function of an utterance, i.e. speech act, which is mapped onto the dialogue act of the DM, and transforms the transcribed utterances into structured representations via deep dependency parsing [3], rule-based graph transduction [5], and ontology design patterns [11]. A frame semantics [10]-oriented knowledge extraction paradigm is followed in the course of which incrementally abstract representations are distilled: 1. *surface-syntactic* → 2. *deep-syntactic* → 3. *predicate-argument* → 4. conceptual, which are translated into OWL knowledge graphs that capture entities and their relations as OWL n-ary relation patterns, and, in particular, as instantiations of DOLCE Ultralite's (DUL) Description and Situation (DnS) patterns. Cf. Figure 4 for the representations 1–4 of the transcription '*I feel sad*'. Its knowledge graph representation is a declarative statement containing an instantiation of the `dul:Situation` class, which interprets the instances of `:CareRecipient` and `:Sad` classes as the experiencer and experienced emotion respectively of the event class `:Feel` instance:

```
:declare a da:Declare ;
   da:containsSemantics :feelCtx1  .
:feelCtx1 a dul:Situation ;
  dul:includes :user1 ;
  dul:includes :sad1 ;
  dul:includesEvent :feel1 ;
```

[2] Essential is also the recognition of prosody as a means to detect the thematic and emphatic patterns in the move of the user [6, 7].

[3] http://www.vocapia.com/

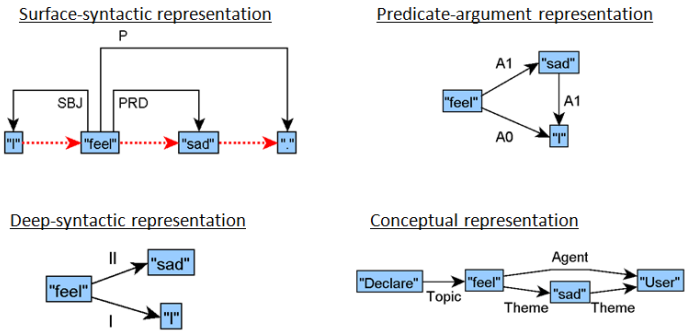

Fig. 4. Example semantic language analysis representations

```
  dul:satisfies :feelDesc1 [a dul:Description] .
:feel1 a :Feel [rdfs:SubClassOf dul:Event] ;
  dul:classifiedBy :Context [rdfs:SubClassOf dul:Concept] .
:sad1 a :Sad [rdfs:subClassOf :Emotion] ;
    dul:classifiedBy :Theme [a dul:Concept] .
:user1 a :CareRecipient [rdfs:SubClassOf dul:Person] ;
    dul:classifiedBy :Experiencer [rdfs:SubClassOf dul:Concept] .
```

Multimodal cues that reflect certain affective states are measured and recognized through the application of sensor technologies, signal processing and recognition techniques. Facial and paralinguistic cues are the most prominent cues. Traditionally, affective face analysis revolved around the recognition of static facial expressions [8, 30]. Nowadays there is a consensus on the need for a dynamic analysis. Commonly, *Action Units* (AUs) from the *Facial Action Coding System* [9] are used as a standard representation [31] . In order to determine facial AUs in a fully automatic manner, we first extract SIFT-based features from sets of automatically detected facial landmarks and then apply a set of independent linear classifiers to associate a probability to each of the targeted AUs. The classifiers are trained following [29], which allows training AU classifiers using datasets with a reduced amount of ground truth (only prototypical facial expressions are needed). Extraction of paralinguistic affective cues is done following [33]. Extracted facial and paralinguistic cues are combined through fusion strategies in order to generate a final prediction. Our work on fusion draws on Lingenfelser's [20] "event-driven" fusion, which is based on [13]. The algorithm does not force decisions throughout considered modalities for each time frame, but instead asynchronously fuses time-sensitive events from any given number of modi. This has the advantage of incorporating temporal alignments between modi and being very flexible with respect to the type and mode of used events. In [20], this algorithm was used to combine the recognition of short-timed laugh (audio) and smile (video) events for a continuous assessment of a user's level of positive valence. For KRISTINA, it is extended to cover the whole valence arousal space, spanned by positive and negative valence and arousal axes.

4.3 Multimodal communication generation

Once the appropriate system action has been determined by the DM, the fission module assigns to the individual mode generation modules the content elements from the OWL graph that are to be expressed by the respective mode. Language generation follows the inverse cascade of processing stages depicted for analysis; see Figure 5 for the successive representations of the system reaction in our running example, namely the suggestion to present to the user news harvested from social media. As generation framework, we use multilingual rule-based [35] and statistical [2] graph transduction modules, which are further adapted to the idiosyncrasies of spoken language. The surface sentence is then spoken by the agent using the CereProc TTS.[4]

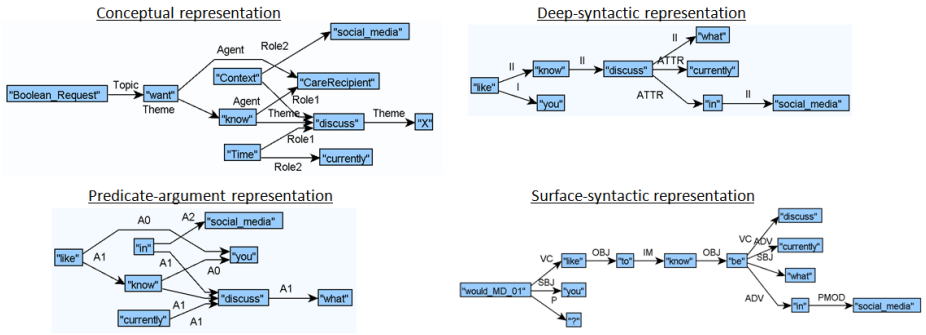

Fig. 5. Example language generation representations

For its non-verbal appearance, KRISTINA is realized as an embodied conversational agent (ECA). The embodiment is realized through a credible virtual character. Credibility (as opposed to realism) implies the believability of the rendering of the agent, avoidance of the trap of the uncanny valley [24], and animation through facial expressions and gestures, when appropriate. Gestures and facial expressions are generated according to the semantics of the message that is to be communicated. Since the generation of facial expressions using tags (smile, surprise, etc.) would limit the possible facial expressions and require a manual design of all possible expressions for each character, we use the valence-arousal representation of emotions [14]; cf., also [17, 18]. Our model can generate and animate facial expressions in the continuous 2D and 3D valence-arousal space by linearly interpolating only five extreme facial poses. Because of its parametric nature, the valence-arousal space can be easily applied to a variety of faces. Using the semantics and other features, gestures are generated, keeping in mind the cultural context of the conversation.

[4] https://www.cereproc.com/

5 Conclusions

We presented the first prototype of a knowledge-centred ECA, which is aimed to conduct socially competent emotive multilingual conversations with individuals in need of advice and support in the context of basic care and healthcare. So far, the agent's conversation skills are restricted to German, Polish, and Spanish; Arabic and Turkish are about to be added. Three different use cases have been setup to validate the progressively increasing functionality of the agent. In the first, it acts as a social companion of elderly with German respectively Turkish background, in the second as an assistant of Polish carers, and in the third as an healthcare adviser of migrants with North African background. Evaluation trials of the 1st prototype have been carried out with users from Germany and Spain with respect to trustworthiness, competence, naturalness of the avatar, friendliness, speech and language understanding and production quality, etc. Cf. the outcome of the questionnaire (on a Likert scale from '1' ("disagree") to '5' ("compeletely agree") on the competence of KRISTINA in Table 1.

Evaluation statement	Likert scale value (SD)
It is clear what KRISTINA wants to communicate	3.23 (± 1.42)
KRISTINA does not provide the right amount of information	2.73 (±1.10)
The conversation with KRISTINA is confusing	2.84 (± 1.27)
KRISTINA behaved as expected	3.0 (± 1.21)
KRISTINA acted on own initiative	3.25 (± 1.29)

Table 1. Outcome of the evaluation of the competence of the 1st prototype

Acknowledgments

The presented work is funded by the European Commission as part of the H2020 Programme, under the contract number 645012–RIA. Many thanks to our colleagues from the University of Tübingen, German Red Cross and semFYC for the definition of the use cases, constant feedback, and evaluation!

References

1. Anderson, K., André, E., Baur, T., Bernardini, S., Chollet, M., Chryssadou, E., Damian, I., Ennis, C., Egges, A., Gebhard, P., Jones, H., Ochs, M., Pelachaud, C., Porayska-Pomsta, K., Rizzo, P., Sabouret, N.: The TARDIS framework: intelligent virtual agents for social coaching in job interviews. In: Reidsma, D., Katayose, H., Nijholt, A. (eds.) ACE, vol. LNCS, 8253, pp. 476–491. Springer, Heidelberg (2013)
2. Ballesteros, M., Bohnet, B., Mille, S., Wanner, L.: Data-driven sentence generation with non-isomorphic trees. In: Proceedings of the 2015 Conference of the NAACL: Human Language Technologies. pp. 387–397. ACL, Denver, Colorado (May–June 2015), http://www.aclweb.org/anthology/N15-1042

3. Ballesteros, M., Bohnet, B., Mille, S., Wanner, L.: Data-driven deep-syntactic dependency parsing. Natural Language Engineering 22(6), 939–974 (2016)
4. Baur, T., Mehlmann, G., Damian, I., Gebhard, P., Lingenfelser, F., Wagner, J., Lugrin, B., André, E.: Context-Aware Automated Analysis and Annotation of Social Human-Agent Interactions. ACM Transactions on Interactive Intelligent Systems 5(2)
5. Bohnet, B., Wanner, L.: Open soucre graph transducer interpreter and grammar development environment. In: Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta (2010)
6. Domínguez, M., Farrús, M., Burga, A., Wanner, L.: Using hierarchical information structure for prosody prediction in content-to-speech application. In: Proceedings of the 8thInternational Conference on Speech Prosody (SP 2016). Boston, MA (2016)
7. Domínguez, M., Farrús, M., Wanner., L.: Combining acoustic and linguistic features in phrase-oriented prosody prediction. In: Proceedings of the 8thInternational Conference on Speech Prosody (SP 2016). Boston, MA (2016)
8. Du, S., Tao, Y., Martinez, A.M.: Compound facial expressions of emotion. Proceedings of the National Academy of Sciences 111(15), E1454–E1462 (2014)
9. Ekman, P., Rosenberg, E.L.: What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA (1997)
10. Fillmore, C.J.: Frame semantics, pp. 111–137. Hanshin Publishing Co., Seoul, South Korea (1982)
11. Gangemi, A.: The Semantic Web – ISWC 2005: 4th International Semantic Web Conference, Galway, Ireland, November 6-10, 2005. Proceedings, chap. Ontology Design Patterns for Semantic Web Content, pp. 262–27 (2005)
12. Gebhard, P., Mehlmann, G.U., Kipp, M.: Visual SceneMaker: A Tool for Authoring Interactive Virtual Characters. Journal of Multimodal User Interfaces: Interacting with Embodied Conversational Agents, Springer-Verlag 6(1-2), 3–11 (2012)
13. Gilroy, S.W., Cavazza, M., Niranen, M., André, E., Vogt, T., Urbain, J., Benayoun, M., Seichter, H., Billinghurst, M.: Pad-based multimodal affective fusion. In: Affective Computing and Intelligent Interaction and Workshops (2009)
14. Gunes, H., Schuller, B.: Categorical and dimensional affect analysis in continuous input: Current trends and future directions. Image and Vision Computing 31(2), 120–136 (2013)
15. Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., von Wilamowitz-Moellendorff, M.: Gumo–the general user model ontology. In: User modeling 2005. Springer, Berlin / Heidelberg (2005)
16. Hofstede, G.H., Hofstede, G.: Culture's consequences: Comparing values, behaviors, institutions and organizations across nations. Sage (2001)
17. Hyde, J., Carter, E.J., Kiesler, S., Hodgins, J.K.: Assessing naturalness and emotional intensity: a perceptual study of animated facial motion. In: Proceedings of the ACM Symposium on Applied Perception. pp. 15–22. ACM (2014)
18. Hyde, J., Carter, E.J., Kiesler, S., Hodgins, J.K.: Using an interactive avatar's facial expressiveness to increase persuasiveness and socialness. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. pp. 1719–1728. ACM (2015)
19. Lamel, L., Gauvain, J.: Speech recognition. In: R, pp. 305–322 (2003)
20. Lingenfelser, F., Wagner, J., André, E., McKeown, G., Curran, W.: An event driven fusion approach for enjoyment recognition in real-time. In: MM. pp. 377–386 (2014)

21. Mehlmann, G., André, E.: Modeling Multimodal Integration with Event Logic Charts. In: Proceedings of the 14th International Conference on Multimodal Interaction. pp. 125–132. ACM, New York, NY, USA (2012)
22. Mehlmann, G., Janowski, K., André, E.: Modeling Grounding for Interactive Social Companions. Journal of Artificial Intelligence: Social Companion Technologies, Springer-Verlag 30(1), 45–52 (2016)
23. Mehlmann, G., Janowski, K., Baur, T., Häring, M., André, E., Gebhard, P.: Exploring a Model of Gaze for Grounding in HRI. In: Proceedings of the 16th International Conference on Multimodal Interaction. pp. 247–254. ACM, New York, NY, USA (2014)
24. Mori, M., MacDorman, K.F., Kageki, N.: The uncanny valley [from the field]. Robotics & Automation Magazine, IEEE 19(2), 98–100 (2012)
25. Motik, B., Cuenca Grau, B., Sattler, U.: Structured objects in owl: Representation and reasoning. In: Proceedings of the 17th international conference on World Wide Web. pp. 555–564. ACM (2008)
26. Ochs, M., Pelachaud, C.: Socially Aware Virtual Characters: The Social Signal of Smiles. IEEE Signal Processing Magazine 30(2), 128–132 (2013)
27. Posner, J., Russell, J., Peterson, B.: The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development and psychopathology. Development and psychopathology 17(3) (2005)
28. Riaño, D., Real, F., Campana, F., Ercolani, S., Annicchiarico, R.: An ontology for the care of the elder at home. In: Proceedings of the 12th Conference on Artificial Intelligence in Medicine: Artificial Intelligence in Medicine. pp. 235–239. AIME '09, Springer-Verlag, Berlin, Heidelberg (2009), http://dx.doi.org/10.1007/978-3-642-02976-9_33
29. Ruiz, A., Van de Weijer, J., Binefa, X.: From emotions to action units with hidden and semi-hidden-task learning. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3703–3711 (2015)
30. Sandbach, G., Zafeiriou, S., Pantic, M., Yin, L.: Static and dynamic 3d facial expression recognition: A comprehensive survey. Image and Vision Computing 30(10), 683–697 (2012)
31. Savran, A., Sankur, B., Bilge, M.T.: Regression- based intensity estimation of facial action units. Image and Vision Computing 30(10), 774–784 (2012)
32. Shaw, R., Troncy, R., Hardman, L.: Lode: Linking open descriptions of events. In: 4th Asian Conference on The Semantic Web. pp. 153–167. Shanghai, China (2009)
33. Wagner, J., Lingenfelser, F., André, E.: Building a robust system for multimodal emotion recognition, pp. 379–419. John Wiley & Sons, Hoboken, NJ (2015)
34. Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., André, E.: The social signal interpretation (SSI) framework–multimodal signal processing and recognition in real-time. In: Proceedings of ACM International Conference on Multimedia (2013)
35. Wanner, L., Bohnet, B., Bouayad-Agha, N., Lareau, F., Nicklaß, D.: MARQUIS: Generation of user-tailored multilingual air quality bulletins. Applied Artificial Intelligence 24(10), 914–952 (2010)
36. Yasavur, U., Lisetti, C., Rishe, N.: Lets talk! speaking virtual counselor offers you a brief intervention. Journal of Multimodal User Interfaces 8(4), 381–398 (2014)
37. Zeng, Z., Pantic, M., Roisman, G., Huang, T.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE transactions on pattern analysis and machine intelligence 31(1), 39–58 (2009)