An Embodied Dialogue System with Personality and Emotions

Stasinos Konstantopoulos

NCSR 'Demokritos', Athens, Greece
konstant@iit.demokritos.gr

Abstract

An enduring challenge in human-computer interaction (HCI) research is the creation of natural and intuitive interfaces. Besides the obvious requirement that such interfaces communicate over modalities such as natural language (especially spoken) and gesturing that are more natural for humans, exhibiting affect and adaptivity have also been identified as important factors to the interface's acceptance by the user. In the work presented here, we propose a novel architecture for affective and multimodal dialogue systems that allows explicit control over the personality traits that we want the system to exhibit. More specifically, we approach personality as a means of synthesising different, and possibly conflicting, adaptivity models into an overall model to be used to drive the interaction components of the system. Furthermore, this synthesis is performed in the presence of domain knowledge, so that domain structure and relations influence the results of the calculation.

1 Introduction

An enduring challenge in human-computer interaction (HCI) research is the creation of natural and intuitive interfaces. Besides the obvious requirement that such interfaces communicate over modalities such as *natural language* (especially spoken) and *gesturing* that are more natural for humans, exhibiting *affect* and *adaptivity* have also been identified as important factors to the interface's acceptance by the user.

We perceive HCI systems as ensembles of interaction modules, each controlling a different interaction modality, and able to modulate their operation depending on external (to the modules themselves) parameters. A central cognitive module deliberates about dialogue acts and orchestrates the interaction modules in order to ensure that such dialogue acts are carried out in a coherent way, keeping uttered content and affect consistent within and across interaction modules.

In this paper we describe work towards this end, carried out in the context of the INDIGO project, and implemented in the form of a *personality module* that complements INDIGO's dialogue manager by calculating parameters related to adaptivity and emotion to be used by the interaction modules in the process of concretely realizing the abstract dialogue-action directives issued by the dialogue manager. This calculation involves the planned act, the user adaptivity model, the system's own goals, but also a machine representation of the *personality* that we want the system to exhibit, so that systems with different personality will react differently even when in the same dialogue state and with the same user or user type.

This is motivated by the fact that, although personality is a characteristically human quality, it has been demonstrated that human users attribute a personality to the computer interfaces they use, regardless of whether one has been explicitly encoded in the system's design (Nass et al., 1995). Furthermore, personality complementarity and similarity are important factors for the acceptance of an interface by a user (Moon and Nass, 1996; Nass and Lee, 2000), so that there is no 'optimal' or 'perfect' system personality, but rather the need to tune system personality to best fit its users.

In the rest of this paper, we will briefly discuss literature on both adaptivity and personality modelling (Section 2), proceed to present the interaction between multimodal dialogue strategies and our personality model (Section 3), and finally conclude (Section 4).

Proceedings of the 2010 Workshop on Companionable Dialogue Systems, ACL 2010, pages 31–36,
Uppsala, Sweden, 15 July 2010.

2 Background

INDIGO in general and our work in particular is, to a large extend, based on work on adaptive natural-language interfaces to databases. The domains of application of these systems have varied from generating personalized encyclopedia entries and museum exhibit descriptions, to supporting the authoring of technical manuals and on-line store catalogues.

2.1 Adaptive HCI

The ILEX system was a major milestone in adaptive *natural language generation* (NLG), emphasising the separation between domain and linguistic resources permitting the *portability* of linguistic resources between domains. ILEX also introduced the notion of a *system agenda* that represents the system's own communicative goals, a significant step in the direction of representing system personality. These system preferences were combined with user preferences and a dynamic *assimilation score* (calculated from interaction history) to estimate a single preference factor for the various facts in the database for the purposes of selecting the content that is to be included in the description of each object (Ó Donnell et al., 2001).

ILEX, however, offered no theory about where interest and importance come from or how to combine them; arbitrary values had to be provided for all objects in the database and the combined preference was derived by multiplying the three factors (importance, interest, and assimilation) regardless of how each object is related to other interesting or important objects in the collection or what other relevant and semantically similar objects have been assimilated.

Building upon ILEX, the M-PIRO system extended user model preferences to influence surface realization besides content selection, so that different surface forms would be generated to realize the same abstract piece of information for different users (Isard et al., 2003). This was achieved by explicitly representing the grammar fragments that could be used to realize different types of facts (properties of the object being described) and then extending the user interests mechanism to also select which grammar fragment is more 'interesting' (or, rather, appropriate) to realize a particular piece of information for a particular user model.

By comparison to ILEX, M-PIRO offered greater flexibility and linguistic variation, as well as language portability by allowing the combination of different grammars with the same domain or user models. On the other hand, the, even rudimentary, ability to combine user and system preferences was dropped and user model authoring became practically unmanageable due the size and complexity of user models.

With the emergence of the Semantic Web, it became obvious that representation technologies such as RDF and OWL offered an opportunity to reduce the authoring effort by operating upon pre-existing OWL ontologies. This motivated the development of the NATURALOWL/ELEON system. NATURALOWL is a template-based NLG engine, explicitly designed for generating natural language descriptions of ontological entities, based on such entities' abstract properties (Galanis and Androutsopoulos, 2007). The ELEON authoring tool (Konstantopoulos et al., 2009) can be used to annotate OWL ontologies with linguistic and content-selection resources and inter-operates with NATURALOWL which can use such annotations to generate descriptions of ontological objects.

2.2 Emotions and personality

Another relevant line of research is centred around *affective interaction* and *intelligent virtual agents*. The main focus here is the modelling and mimicking of the various affective markers that people use when they communicate, aiming at more natural and seamless human-computer interaction.

Such affective systems are modulated by *personality representations* varying from fully-blown cognitive architectures (Vankov et al., 2008) to relatively simpler personality models. The *OCEAN* or *Big Five* model, in particular, a standard framework in psychology (Norman, 1963; Costa and McCrae, 1992), is used to represent personality in a variety of virtual agents and avatars capable for multi-modal communication acts such as speech and facial expressions (Strauss and Kipp, 2008; Kasap et al., 2009). Such systems are typically rich in visual expression, but lack sophistication in natural language generation, knowledge representation and dialogue structure.

The PERSONAGE and INDIGO systems, on the other hand, move in the area between these systems and the database-access systems discussed above: PERSONAGE develops a comprehensive

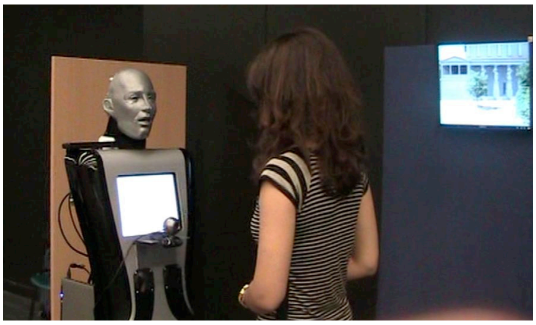

Figure 1: An INDIGO robot interacting with *Hellenic Cosmos* personnel during preliminary trials, September 2009.

theory of using OCEAN parameters to control natural language interaction from lexical choice to syntax, pragmatics, and planning, but is restricted to text generation and no other communication modalities are covered (Mairesse and Walker, 2007). The INDIGO dialogue system emphasises multi-modality as it is embodied in a robot capable of multi-modal interaction. INDIGO uses OCEAN to combine a separate user model and system profile into a single parameter set used to parametrize a number of interaction components, such as a virtual avatar capable of displaying emotions, the NLG engine, the text-to-speech engine, the dialogue manager, etc.

3 A dialogue system with personality

The INDIGO system has been fielded at the *Hellenic Cosmos* cultural centre,[1] where it provides personalized tours with historical, architectural, and cultural information about the buildings of the Ancient Agora of Athens (Figure 1).

The *dialogue manager* (DM, Matheson et al., 2009), implemented using TrindiKit,[2] assumes the *information-state and update* approach to dialogue management (Traum and Larsson, 2003). The information state stores information such as dialogue history and current robot position. Input from the sensors (ASR, vision, laser tracker, and touchscreen) is processed by update rules which heuristically fuse multimodal (and possibly contradicting) sensory input and implement generic (i.e., domain and personality-independent) dialogue strategies. These strategies deliberate about the next action that the robot will take, such as moving to a different section of the exhibition, offering a menu of choices, or describing an item.

One notable strategy implemented in the DM is the *Move On Related* strategy (Bohus and Rudnicky, 2008), the system's fallback when user input cannot be confidently recognized even after fusing all input modalities. In such situations, DM uses the combined preference factors to choose the most preferred exhibit within the ontological class that is the current focus of the discourse. If there is an instance in this class with a clear preference, DM assumes this as the user response; if, on the other hand, there is no instance with significantly higher preference than the rest, DM prompts the user to repeat their answer or use the touchscreen.

The other notable, and widely used, strategy is the one that drives the two loops shown in Figure 2, in response to a user request for content: one pertaining to dynamically realizing a personalized description of an object of the domain ontology and one pertaining to updating the system's emotion and mood.

3.1 Content selection and realization loop

Once the DM has resolved that the next robot action will be the description of a domain ontology object, the personality-driven preferences are used to select which properties of this object will be included in the description. These preferences are calculated taking into account a combined user-system preference (Konstantopoulos et al., 2008) as well as a dynamic *assimilation score*, calculated from interaction history, which balances between the gratuitous and tiring repetition of high-preference material and simply rotating through the list of properties of an object.

The chosen content is then used by the NATURALOWL NLG engine (Galanis and Androutsopoulos, 2007) to plan and realize a personalized textual description of the object. Besides selecting what to include in a description, preference is used by NATURALOWL to annotate the generated text with directives, such as *emphasis*, for the text-to-speech effector that drives the robot's speakers.

The combined user-system preference stems from associating domain objects with content-selection parameters, using an representation developed for NATURALOWL and extended in INDIGO to provide for representing not only user models but also *system profiles* that establish the system's own goals and preferences (Konstan-

[1] See also `http://www.hellenic-cosmos.gr`
[2] See `http://sourceforge.net/projects/trindikit/`

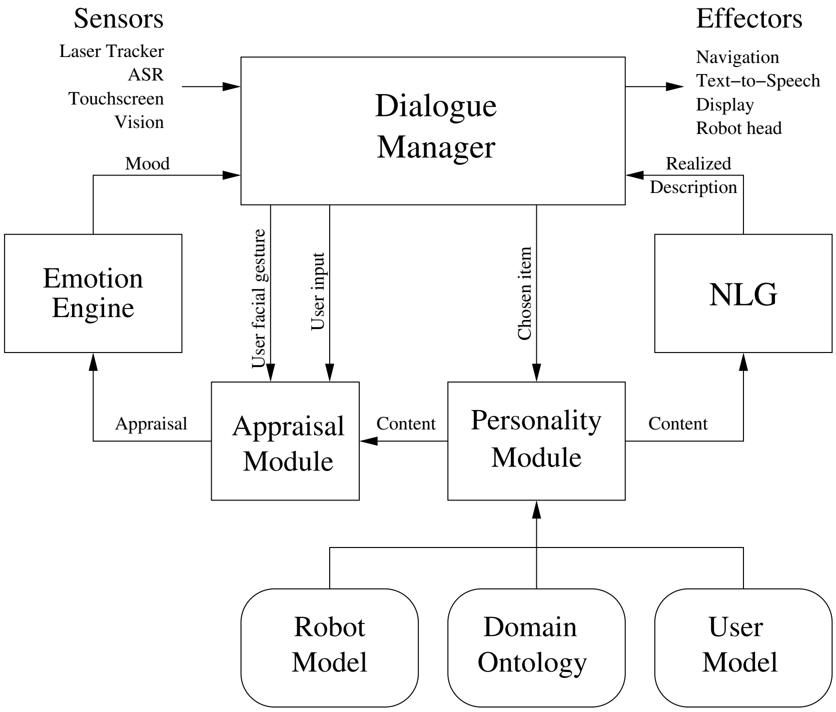

Figure 2: Overall architecture of the dialogue system.

topoulos et al., 2009).

Emotional and, in general, behavioural variation among different instantiations of the system is achieved through *synthetic personality models* that assert different points of balance between the (potentially) conflicting user and system preferences. What is of particular importance is that the the combined user-system preference is not estimated in isolation for each and every ontological object as was the case in ILEX, but by axiomatizing how preference is 'transferred' between domain objects based on their semantic relations. This is achieved by defining personality in terms of logic clauses that link the preferences of an object not only to its user and system preferences, but also to those of objects it semantically relates with.

3.2 Emotional appraisal and update loop

The system emotionally appraises user actions as well as its own actions. With respect to its own actions, the preference factors for the properties selected to describe an object reflect the robot's being excited or bored to discuss the current subject.

Appraisal of user actions stems from vision and speech analysis to reflect the impact of the *manner* of what the user said. More specifically, facial gesture recognition is used to detect emotional signs (such as smiling) besides detecting affirmative and negative nods and similar signs that are fused with the results of speech recognition.

As user utterances are mostly short and incomplete answers to questions such as 'Would you like to hear more about this monument?' or 'Which monument would you like me to talk about?' we cannot detect emotion based on linguistic meaning or syntactic structure, but rather concentrate on extracting useful prosodic and linguistic features such the length of the last syllable in an utterance or whether the first word of the utterance is an *wh*-word.[3] Although these features are not by themselves indicative of emotion, they are indicative of prosody and their combination with segmental features (referring to the acoustic form) extracted directly from the speech signal was shown to improve emotion estimation.

Emotional appraisal is used by an *emotion simulator* (Kasap et al., 2009) that uses the system's personality traits (OCEAN vector) to model how dialogue acts affect the system's emotional state. This emotion simulator updates the system's internal short-term *emotional state* and long-term *mood* by applying an update function on the current state and the *emotional appraisal* of each dialogue act. The OCEAN parameters act as parameters of the update function, so that, for example,

[3] Where, what, who, etc.

neuroticism (i.e., 'tendency to distress') makes the update function tend towards negative emotions, whereas agreeableness (i.e., 'sympathetic') makes it more directly reflect the user's emotions.

The speech synthesiser and the robot's animatronic head reflect emotional state as voice modulations and facial expressions, whereas mood is taken into account by the DM when deliberating about the robot's next dialogue action.

4 Conclusions

In this paper we have approached personality as a means of synthesising different, and possibly conflicting, adaptivity models into an overall model to be used to drive the interaction components of the system. Furthermore, this synthesis is performed in the presence of domain knowledge, so that domain structure and relations influence the results of the calculation.

We thusly explore the *self vs. other* aspect of personality modelling, theoretically interesting but also practically important as we cleanly separate adaptivity and profiling data that refers the system from that which refers to the user. This follows up on the tradition of the line of systems stemming from ILEX, where increasingly separable models (domain vs. NLG resources, the latter later broken down between linguistic and adaptivity resources) have allowed for such hard-to-create resources to be re-used.

Acknowledgements

The work described in this paper was funded by the FP6-IST project INDIGO, that developed and advanced human-robot interaction technology, enabling robots to perceive natural human behaviour and act in ways that are more familiar to humans. To achieve its goals, INDIGO advanced various technologies, which it integrated in a robotic platform.

More details are availble from the project website, `http://www.ics.forth.gr/indigo`

References

Dan Bohus and Alex Rudnicky. 2008. Sorry, I didn't catch that. In Laila Dybkjær and Wolfgang Minker, editors, *Recent Trends in Discourse and Dialogue*, volume 39 of *Text, Speech and Language Technology*, chapter 6, pages 123–154. Springer Netherlands.

P. T. Costa and R. R. McCrae. 1992. Normal personality assessment in clinical practice: The NEO personality inventory. *Psychological Assessment*, 4(5–13).

Dimitris Galanis and Ion Androutsopoulos. 2007. Generating multilingual descriptions from linguistically annotated OWL ontologies: the NaturalOWL system. In *Proceedings of the 11th European Workshop on Natural Language Generation (ENLG 2007), Schloss Dagstuhl, Germany*, pages 143–146.

Amy Isard, Jon Oberlander, Ion Androutsopoulos, and Colin Matheson. 2003. Speaking the users' languages. *IEEE Intelligent Systems*, 18(1):40–45.

Zerrin Kasap, Maher Ben Moussa, Parag Chaudhuri, and Nadia Magnenat-Thalmann. 2009. Making them remember: Emotional virtual characters with memory. In Tiffany Barnes, L. Miguel Encarnção, and Chris Shaw, editors, *Serious Games, Special Issue of IEEE Computer Graphics and Applications*. IEEE.

Stasinos Konstantopoulos, Vangelis Karkaletsis, and Dimitris Bilidas. 2009. An intelligent authoring environment for abstract semantic representations of cultural object descriptions. In Lars Borin and Piroska Lendvai, editors, *Proceedings of EACL-09 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education* (LaTeCH-SHELT&R 2009), Athens, 30 Mar 2009, pages 10–17.

Stasinos Konstantopoulos, Vangelis Karkaletsis, and Colin Matheson. 2008. Robot personality: Representation and externalization. In *Proceedings of ECAI-08 Workshop on Computational Aspects of Affective and Emotional Interaction (CAFFEi 2008), Patras, Greece, July 21st, 2008*, pages 5–13.

François Mairesse and Marilyn Walker. 2007. PERSONAGE: Personality generation for dialogue. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*. Prague.

Colin Matheson, Amy Isard, Jon Oberlander, Stasinos Konstantopoulos, and Vangelis Karkaletsis. 2009. Multimodal human-robot dialogue management. INDIGO Deliverable 4.1 (public).

Youngme Moon and Clifford Nass. 1996. Adaptive agents and personality change: complementarity versus similarity as forms of adaptation. In *Proceedings SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, BC, Canada, 1996. SIG on Computer-Human Interaction, ACM, New York, U.S.A.

Clifford Nass and Kwan Min Lee. 2000. Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In *Proceedings SIGCHI Conference on Human factors in Computing Systems*, The Hague, 2000. SIG on Computer-Human Interaction, ACM, New York, U.S.A.

Clifford Nass, Youngme Moon, B. Fogg, and B. Reeves. 1995. Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43:223–239.

W. T. Norman. 1963. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality rating. *Journal of Abnormal and Social Psychology*, 66:574–583.

Michael Ó Donnell, Chris Mellish, Jon Oberlander, and A. Knott. 2001. ILEX: an architecture for a dynamic hypertext generation system. *Natural Language Engineering*, 7(3):225–250.

Martin Strauss and Michael Kipp. 2008. ERIC: a generic rule-based framework for an affective embodied commentary agent. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems (AAMAS 08), Estoril, Portugal, 2008*, pages 97–104.

David Traum and Steffan Larsson. 2003. The information state approach to dialogue management. In Jan van Kuppevelt and Ronnie Smith, editors, *Current and New Directions in Discourse and Dialogue*. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Ivan Vankov, Kiril Kiryazov, and Maurice Grinberg. 2008. Introducing emotions in an analogy-making model. In *Proceedings of 30th Annual Meeting of the Cognitive Science Society (CogSci 2008), Washington D.C.*