On Intuitive Dialogue-based Communication and Instinctive Dialogue Initiative

Daniel Sonntag

German Research Center for Artificial Intelligence
66123 Saarbrücken, Germany
sonntag@dfki.de

Introduction

AI helps to solve fundamental problems of human computer interaction (HCI) technology, such as dialogue-based communication with machines. Who wouldn't like to speak freely to computers and ask questions which could be answered in real-time with the help of search engines on the World Wide Web or other information repositories? Eventually, the role of dialogue systems may shift from merely performance enhancers (e.g., voice input is fast and convenient on mobile devices) toward guides, educational tutors, or adaptable interfaces in ambient intelligence environments. Intelligent user interfaces (IUIs) may be understood as human-machine interfaces that aim to improve the efficiency, effectiveness, and naturalness of human-machine interaction.

In order to implement these properties, explicit models of the discourse of the interaction, the available information material, the domain of interest, the task, and/or models of a user or user group have to be employed. But that's not everything. Dialogue-based interaction technology, and HCIs in general, still have plenty of room for improvements. For example, dialogue systems are very limited to user adaptation or the adaptation to special dialogue situations. In fact, humans adapt their dialogue behaviour over time according to their dialogue partners' knowledge, attitude, and competence. This is possible because humans' abilities further include (1) the emotions that are expressed and perceived in natural human-human communication, (2) the instinctive actions and reactions a human dialogue participant performs, and (3) the metacognitive and self-reflective abilities of a human dialogue participant to cleverly reason about the actions she or he takes.

Interestingly, as [1] points out, emotions can only be inferred from context, self-report, and the expressive (dialogue) behaviour. This points to human-centred human-computer interaction strategies, in order to enable computers to unobstrusively respond to the user-perceived content. This, in turn, corresponds our provided definition of IUIs. We hypothesise that the underlying maxims of conversation and the resulting (multimodal) dialogue constraints may very much be related to instinctive computing as *instinctive dialogue initiative*, for which we will provide evidence. We attempt to shed light on the relationship between instinctive computing and state-of-the-art multimodal dialogue systems in order to overcome the limitations of contemporary HCI technology.

Limited HCI Technology and Dialogue-based Solutions

Interaction technology and HCIs are extremely popular. However, the technology, especially the effective retrieval of information while using advanced user interfaces, such as multmodal dialogue, is still in a stage of infancy. We are often working with systems that use canned dialogue segments, hardwired interaction sequences, no inference services, and very limited adaptation possibilities (cf. figure 1, left). We can work against the limitations of current multimedia and HCI technology by exploiting dialogue systems with metacognitive abilities and interaction agents that can simulate instincts. We distinguish foraging, vigilance, reproduction, intution, and learning as the human basic instincts (also cf. [3]). (Foraging and reproduction have no embodiment in contemporary AI for interaction technology and HCIs.) Thereby, dialogue systems should become more prevalent in modelling the learning aspect whereas the instinctive agent should become more prevalent in implementing intuition. Vigilance is somewhere in between as the process of self-adaptation or innate (apparently unlearned) behaviour patterns (cf. figure 1, right).

We ask the question which developments in multimodal dialogue systems allow for advancements in computing instincts. One could argue that cognitive instincts and (meta-) cognitive dialogue strategies use the same class of actual sensory input. Likewise, we are interested in how instincts relate to dialogue constraints, and the attempts to convey them, in order to make HCIs more intelligent. Basically, the advancement in multimodal dialogue systems we proposed in [2] is introspection for meta dialogue which should now be used for modelling instinctive dialogue initiative. There are several ways to advantageously combine these methods.

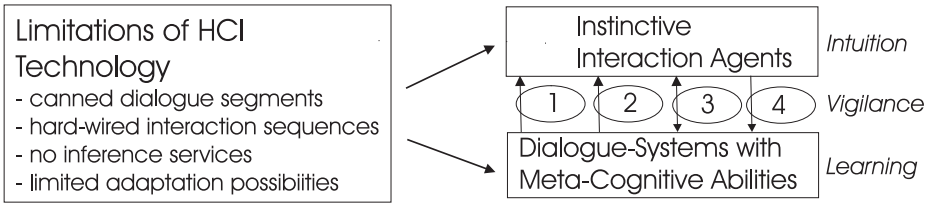

Fig. 1. Limitations of HCI Technology. Combined methods to overcome the limitations include (1) obeying dialogue constraints, (2) modelling self-reflection and adaptation, (3) using sensory methods, and (4) implementing intuition.

(1) Obeying Dialogue Constraints. Dialogue constraints subsume four constraint types: linguistic constraints (e.g., correct case and number generation), correct dialogue acts as system responses (cf. adjacency pairs, for example), timing constraints, and constraints on the information content itself (e.g., information to be presented should generally follow Grice's maxims and the users' presumptions

about utterances; information should be available in an appropriate quantity, for example). Also see [4] for a list of social discourse obligations. In the context of question answering and our definition of IUIs as being a particularly efficient, effective, and natural implementation of human-machine interaction, we identified the following four system initiative constraints: (1) retain the user by reporting on the question processing status, (2) informing the user about the probability of query success, (3) informing the user as to why the current HCI process is due to fail, (4) balancing the user and system initiative. Furthermore, we propose extending these key aspects of system initiative to a notion of *instinctive dialogue initiative.*

(2) Modelling Self-Reflection and Adaptation. Humans use metacognition to monitor and control themselves, to choose goals, to assess their progress, and to adopt new strategies for achieving goals. Psychological literature provides a wide array of influences on metacognition that emphasises cognitive self-monitoring, self-reflection, and the importance of explicit representations for adapting one's behaviour. We use a two-level structure to implement the self-reflection and adaptation mechanism, whereby the cognitive processes are split into two interrelated levels: The meta-level (metacognition) contains a dynamic model of the object-level (cognition); the two dominant relations between the levels are called control and monitoring [5]. In dialogue systems, the dialogue manager, which contains all action rules, observes the dialogue progress (monitoring), builds machine learning models about failure and success cases, and updates its internal reasoning model for taking actions. Our experiments are in the context of dialogue-based question answering. A dialogue shell, with state-of-the-art natural language processing methods combined in a Semantic Web framework, takes care of the task-based dialogue [6, 7]. We were able to predict empty results, answer times, and classify queries for the probability of success according to query features and specific access/quality properties of the answer services in a changing environment. For example, as response to a question, we can initiate a system reaction that automatically informs the user "An appropriate answer is not in my knowledge base; I will search the Internet for a suitable answer;" or "I need some time, empty results are not expected, but the results won't be entirely certain."

(3) Using Sensory Methods. Multimodal interaction scenarios and user interfaces may comprise a lot of different sensory inputs. For example, speech can be recorded by a bluetooth micro and sent to an automatic speech recogniser; camera signals can be used to capture facial expressions; the user state can be extracted using biosignal input, in order to interpret the current stress level of the user. The latter point corresponds to an instinctive preliminary estimate of a dialogue participant's emotional state. In addition, several other sensory methods exist that can be used for a dialogue's situational and discourse context—all of which can be seen as an instinctive sensory input: First, the attention detection using on-focus/off-focus. If you are addressed with the eyes in, e.g., a multi-party conversation, you are more vigilant that you will be the next to

take over the dialogue initiative. (This is similar to eye-tracker functionality predominantly used in usability studies to learn how to reduce the cognitive load.) Second, with anthropocentric interaction design and models, we seek to build input devices that can be intuitively used. We recognised that the thumb plays a significant role in modern society—becoming humans' dominant haptic interactor. This development should be reflected in the interface design for future HCIs. Whether society-based interaction habits (e.g., you subconsciously decide to press a doorbell with your thumb) can be called an instinctive way of interaction, is just one aspect of the debate about the relationship between intuition and instincts (also cf. last paragraph).

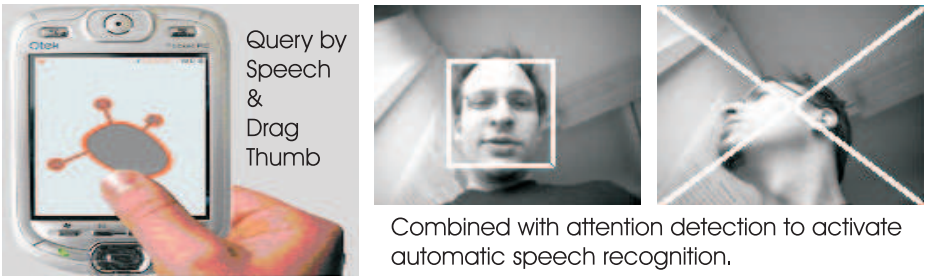

Fig. 2. Anthropocentric thumb sensory input on mobile touchscreen (left) and two still images illustrating the function of the OnView/OffView (right).

(4) Implementing Intuition. According to paragraph (3), cognitive instincts and (meta-) cognitive dialogue strategies use the same class of actual sensory input. When it comes to implementing intuition, we expect the instinctive interaction agent to deliver the appropriate action rules which are then included in the dialogue decision process by following the same procedure as with the adaptation models. Although intuition is widely understood as a non-perceptual input to the decision process (*In philosophy, the power of obtaining knowledge that cannot be acquired either by inference or observation, by reason or experience.* (Encyclopdia Britannica)), only one single dialogue manager update class is necessary for its internal reasoning model for taking actions. Nonetheless, as *[Instinct is] an inborn impulse or motivation to action typically performed in response to specific external stimuli* (Encyclopdia Britannica), the action model could be incorporated into a meta-control-based dialogue manager (cf. the two-level structure). Hence, a useful and cooperative question answering dialogue in natural language would not only combine different topics, heterogeneous information sources, and user feedback, but also intuitive meta dialogue—initiated by an instinctive interaction agent.

References

1. Cohn, J.F.: Foundations of human computing: Facial expression and emotion. [8] 1–16
2. Sonntag, D.: Introspection and adaptable model integration for dialogue-based question answering (forthcoming). In: Proceedings of the Twenty-first International Joint Conferences on Artificial Intelligence (IJCAI). (2009)
3. Cai, Y.: Instinctive computing. [8] 17–46
4. Kaizer, S., Bunt, H.: Multidimensional dialogue management. In: Proceedings of the 7th SigDial Workshop on Discourse and Dialogue, Sydney, Australia (July 2006)
5. Nelson, T.O., Narens, L.: Metamemory: A theoretical framework and new findings. In: G. H. Bower (Ed.) The Psychology of Learning and Motivation: Advances in Research and Theory. Volume 26. Academic Press (1990) 125–169
6. Reithinger, N., Sonntag, D.: An integration framework for a mobile multimodal dialogue system accessing the Semantic Web. In: Proceedings of INTERSPEECH, Lisbon, Portugal (2005) 841–844
7. Sonntag, D., Engel, R., Herzog, G., Pfalzgraf, A., Pfleger, N., Romanelli, M., Reithinger, N. [8] 272–295
8. Huang, T.S., Nijholt, A., Pantic, M., Pentland, A., eds.: Artifical Intelligence for Human Computing, ICMI 2006 and IJCAI 2007 International Workshops, Banff, Canada, November 3, 2006, Hyderabad, India, January 6, 2007, Revised Seleced and Invited Papers. In Huang, T.S., Nijholt, A., Pantic, M., Pentland, A., eds.: Artifical Intelligence for Human Computing. Volume 4451 of Lecture Notes in Computer Science., Springer (2007)