Künstl Intell
DOI 10.1007/s13218-012-0169-4

FACHBEITRAG

Collaborative Multimodality

Daniel Sonntag

Received: 9 August 2011 / Accepted: 19 January 2012

Abstract This essay is a personal reflection from an Artificial Intelligence (AI) perspective on the term HCI. Especially for the transfer of AI-based HCI into industrial environments, we survey existing approaches and examine how AI helps to solve fundamental problems of HCI technology. The user and the system must have a collaborative goal. The concept of *collaborative multimodality* could serve as the missing link between traditional HCI and intuitive human-centred designs in the form of, e.g., natural language interfaces or intelligent environments. Examples are provided in the medical imaging domain.

Keywords AI methods · Multimodal interaction · Dialogue systems · Collaboration

1 Introduction

The term Human-Computer Interaction (HCI) confuses some researchers and practitioners. Many think of HCI as including diverse areas of traditional graphical and web user interfaces, others rather think of new multimodal input and output devices, tangible user interfaces, virtual and augmented reality, intelligent environments, and/or interfaces in the ubiquitous computing paradigm. Whereas the supporters of the traditional HCI view have a strong motivation and justification in the desktop-based ubiquitousness of traditional computer terminals with computer screens, psychological analysis background, and integral evaluation methods, the new AI-based technologies can impress with intuitive human-centred designs. (It should be noted that human-centred designs do not necessarily improve the usability of an HCI, especially in industrial environments.)

HCI is the business of designing user interfaces that people can work well with. Hence, it is an area of research, design, and application, which combines all the aforementioned diverse areas. There is a great variety in these partly overlapping areas which are all involved in this business.

In this article, the first goal is to give an overview of the AI-based HCI techniques which include multimodal interaction. We think that in the future, AI will have a great influence on multimodal interaction systems. Therefore, we will begin by articulating the key issues of the concept of multimodal interfaces in the sense of a combination of traditional, screen-based HCI techniques and interfaces with new (mobile) multimodal input and output devices and interaction forms. Goebel and Williams have commented on the goal to stitch together the breadth of disciplines impinging on AI [8]; following their idea, we try to stitch together the breadth of disciplines impinging on AI-based multimodal HCI. For this purpose, we will introduce the notion of *collaborative multimodality*, which could serve as the missing link between traditional HCI and intuitive human-centred designs.

The second goal is to give a summary of the various different approaches taken by ourselves and other participants in the research field of multimodal dialogue-based HCI for prototyping industry-relevant applications of intelligent user interfaces (IUIs). We think that collaborative multimodality as introduced here represents one of the major usability requirements. AI methods such as sensory input interpretation and explicit models of the discourse of the interaction and the domain of interest are employed to create an argument in favour of the hypothesis that the emergence of a complex, collaborative multimodal behaviour is what best describes an intelligent user interface. To prove this,

D. Sonntag (✉)
German Research Center for AI (DFKI), Stuhlsatzenhausweg 3, 66123 Saarbruecken, Germany
e-mail: sonntag@dfki.de

Published online: 02 February 2012

we provide some intermingling criteria for categorising an interface as an intelligent one to achieve collaborative multimodality. Our working hypothesis is that multimodal interaction provides the best background for showing intelligence in user interfaces—either by advanced analytical methods for understanding multiple sensory input modalities or by the emergence of a complex multimodal behaviour which, if performed by a human, would be deemed intelligent.

2 Multimodal Human-Computer Interaction

A modality is a path of communication between the human and the computer. The technical definition speaks of a sensor or device through which the computer (or human) can receive the input from the human (or computer), or send a message. Seeing or vision modality, and hearing or auditory modality are the prominent modalities. Thereby, multiple modes are employed, for example traditional keyboard and mouse input/output, speech, pen, touch, gestures, and gaze. Multimodal systems which combine several modalities allow for mutual disambiguation of input modalities. Pointing gestures on objects on a touchscreen, for example, can be more reliable than recognising and understanding the spoken name of an object.

Multimodal dialogue systems allow dialogical inputs and outputs in more than just one modality and go beyond the capabilities of text-based Internet search engines or speech dialogue systems. Depending on the specific context, the best input and output modalities can be selected and combined. They belong to the most advanced intelligent user interfaces [17] in comparison to text-based search engines, for example.

In general, the interaction between the user and the multimodal dialogue system can either be user-initiative, system-initiative, or mixed-initiative. In user-initiative systems, the user has to utter a command before the system starts processing. Hence, the focus is more or less on the correct interpretation of the user's utterance; the new user intention, therefore, drives the behaviour of the dialogue system. A typical example can be found in route-planning in the transportation domain. For example, User: "I need to get to London in the afternoon"—System: "Ok. Query understood." The system can react by taking initiative and gathering the necessary information for the decision which mode of transportation would be preferable. In system-initiative systems, the user and the system must have a collaborative goal, and after initialisation (the user question), the system basically asks for missing information to achieve this goal. In [12], mixed-initiative is defined as: "[...] the phrase to refer broadly to methods that explicitly support an efficient, natural interleaving of contributions by users and automated services aimed at converging on solutions to problems." This basically means that the sub-goals and commitments have to come from both parties and be cleverly fulfilled and/or negotiated.

3 Collaborative Multimodality

The concept of *collaborative multimodality* could serve as the missing link between traditional HCI and intuitive human-centred designs in the form of, e.g., tangible user interfaces or intelligent environments in industrial settings. Previous research in dialogue systems, see for example [3], has emphasised the important relationship between co-operation and multimodal communication. Especially in industrial application domains of collaborative multimodality, the extended concept of *task-based* co-operation should be one of the major driving forces. That means, first, the dialogue is not, e.g., a chat about the weather conditions or the like, but rather a task such as booking flights, answering specific questions about a particular domain (e.g., medical conditions), or helping a user set up and program a media recorder [27] or the iTunes Store; and second, the participants must have a common interaction goal; collaborative discourse theory can be used to bring together human-machine collaboration and multimodal dialogue. An important side-condition of multimodal interaction principles based on discourse theory is that dialogue utterances are also treated as actions of the multimodal dialogue system manager.

In general, when humans converse with each other, they utilise many input and output modalities in order to interact. These include gestures or mimicry (including drawings or written language, for example), which belong to non-verbal communication. The verbal communication mode is the spoken language. Some modes of communication are more efficient or effective for certain tasks or contexts. For example, a mobile user interface could be addressed by spoken language in contexts where someone is already busy with his hands and eyes (for example, while driving) or simply to avoid tedious text input modes on mobile devices such as smartphones. AI techniques [28] can be used to model this complex interaction behaviour and Machine Learning (e.g., see [20]) plays a significant role in the modelling of discourse information, collaborative goals, and content information over time. We are almost certain that multimodal dialogue-based communication with machines will become one of the most influential AI applications of the future. Who wouldn't like to speak freely to computers and ask questions about clicked or pointed items? At the point of writing this text, Apple shipped the first voice speech input software, Siri, which takes the application context (e.g., email client or web search) into account, to the iPhones.

Only three years ago, multimodal interfaces for coherent dialogue were among the main achievements of the SmartWeb system (funded by the German Federal Ministry of Education and Research with grants totalling 14 million euros). In SmartWeb, questions and commands are (additionally) interpreted according to the context of the previous conversation (Apple has just begun to assimilate the basic technologies, for example, the disambiguation of a context is not yet integrated). We will take this multimodal dialogue infrastructure as an example of how applications like Siri can be implemented (cf. the range and similarity of voice applications to the prior research projects). The full list of collaborative multimodality principles can be summarised as follows:

1. HCI systems must not artificially separate dialogue and action;
2. HCI dialogues should be planned with modality independent systems;
3. HCI systems must be multimodal and use AI technology;
4. HCI systems must employ mixed initiative dialogues.

3.1 Multimodal Dialogue Infrastructures

We learned some lessons which we use as guidelines in the development of basic architectures and software infrastructures for multimodal dialogue systems. In earlier projects [25, 34, 35] we integrated different components into multimodal interaction systems. Hub-and-spoke dialogue frameworks played a major role [26]. We also learned some lessons which we use as guidelines in the development of *semantic* dialogue systems [23, 32]; over the last years, we have adhered strictly to the developed rule "no presentation without representation." The idea is to implement a generic, and semantic, dialogue shell that can be configured for and applied to domain-specific dialogue applications, thereby "planning" everything with modality-independent data structures before a specific input or output modality is used. In this way, the above-mentioned artificial separation of dialogue and action into distinct UI modes does not occur. This principle has already been successfully applied in multimodal dialogue demonstrators to pave the way towards symmetric multimodality for dialogue systems in which all input modes (e.g., speech, gesture and facial expression) are also available for output, and vice versa [35]. The assumption of [4] has not changed—users require more effective and efficient means of interaction with increasingly complex information and new interactive devices (Fig. 1 shows the related scientific fields and enabling technologies). Several studies, e.g., [5], have shown that the study of HCI only through GUIs and the study of natural dialogue are two separate fields. Dialogue-based communication serves well as an inspiration for designing multimodal dialogue demonstrators.

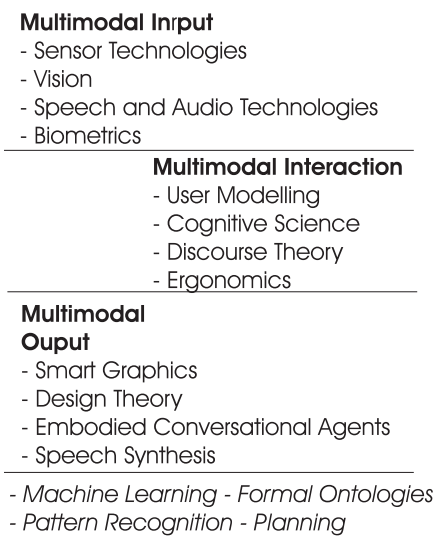

Fig. 1 Scientific fields and enabling technologies [4]

On the technical basis, all messages transferred between internal and external components are now based on RDF data structures (http://www.w3.org/RDF/) which are modelled in a discourse ontology (also cf. [6, 11, 30]). Likewise, the authoring tool *Disco for Games* supports the creation of computer games in which dialogue and action are integrated without the need for changing individual input or output modes [10].

As a result of our developed rule of "no presentation without representation," our systems for industrial dissemination have the following four main properties: (1) multimodality of user interaction, (2) ontological representation of interaction structures and queries, (3) ontological representation of the HCI, and (4) encapsulation of the dialogue proper from the rest of the application.[1] These properties correlate with the first three of our HCI principles for collaborative multimodality.

On the other hand, intelligent AI systems for HCI that involve intelligent algorithms for dialogue processing and interaction management must be judged for their suitability in industrial environments. As a matter of fact, human-centred designs do not necessarily improve the usability of an HCI in a specific application context, especially in industrial environments. This fact represents one of the limitations of multimodal dialogue technology in general, which can be addressed, but not be solely solved by improvements of current AI techniques.

One major concern which we observed in the development process for industrial applications over the last years is that the incorporation of AI technologies such as complex natural language understanding components (e.g., head-driven phrase structure grammar (HPSG) based speech

[1] A comprehensive overview of ontology-based dialogue processing and the systematic realisation of these properties can be found in [30], pp. 71–131.

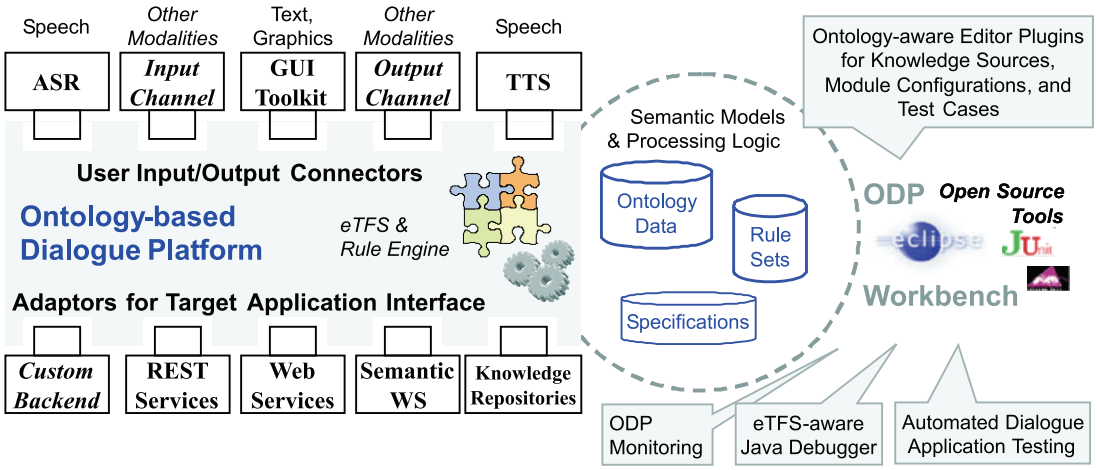

Fig. 2 Overall design of the ontology-based dialogue infrastructure ODP [33]

understanding) and open-domain question answering functionality can unintentionally diminish a dialogue system's usability. This is because negative side-effects such as diminished predictability of what the system is doing at the moment and lost controllability of the internal dialogue processes (e.g., a question answering process) occur more often when AI components are involved. This tendency gives rise to new requirements for usability to account for the special demands introduced by the use of AI.[2]

Also, the predictability would decrease when interfaces are allowed to adapt to a specific user model automatically. Nonetheless, considerable process has been made, for example in automatic graphical user interface generation, where systems generate personalised interfaces that are adapted to the individual motor capabilities of users (assistive technologies for persons with disabilities, i.e., motor impairments) [7].

3.2 Ontology-Based Dialogue Processing (ODP)

Our ODP workbench (Fig. 2) builds upon the industry standard Eclipse and also integrates other established open source software development tools to support dialogue application development, automated testing, and interactive debugging. A distinguishing feature of the toolbox is the built-in support for eTFS (extended Typed Feature Structures), the optimised ODP internal data representation for RDF-based knowledge structures. This enables ontology-aware tools for the knowledge engineer and application developer. Likewise, a fundamental AI field is treated in detail: knowledge representation for industry-relevant interaction systems [33]. An extensively different challenge would be, for example, how intelligent HCIs can ask for user input to improve their reasoning processes [1], additionally based on ontological knowledge representation in the context of research on active learning systems. Figure 2 also shows the ontology components the user works with. The graphical user interfaces (GUIs) for editing ontologies, speech recognition grammars, and interaction rules are implemented as Eclipse plugins in the ODP workbench using the open source toolkit JUnit. The results of the interactive processes where the dialogue engineers are involved, are stored in RDF repositories as ontology data, domain-dependent specifications, and ontology-based rules sets for interaction and input interpretation rules of the specific application domain. We will get back to the architectural issue in the context of modelling self-reflection and adaptation to address and overcome some of the current limitations. New ways to achieve collaborative multimodality include, amongst others, the modelling a dialogue system's initiative to ask for user input to improve the internal reasoning processes (cf. fourth principle).

3.3 Medical Application Example

Clinical radiologists skim many image series and thousands of pictures in a minute's time. Although it is widely reductive to put it this way, a (senior) radiologist has three main goals: (1) access the images and image (region) annotations, (2) complete them, and (3) refine existing annotations. These tasks can best be fulfilled while using a multimodal dialogue system.

The semantic dialogue system, here our ODP system, should be used to ask questions about the image annotations while engaging the clinician in a natural speech dialogue. With the incorporation of higher level knowledge represented in ontologies, different semantic views of the same medical images (such as structural, functional, and disease aspects) can be explicitly stated, integrated, and asked for. This is the essential part of the knowledge acquisition process, and the dialogue system is the knowledge acquisition tool. Two aspects are implemented. First the inspection of and navigation through the patient's data, and second, the

[2]For the identification of these usability issues, the binocular view (AI and HCI) of interactive intelligent systems might be interesting to the reader (discussed in detail in [14]).

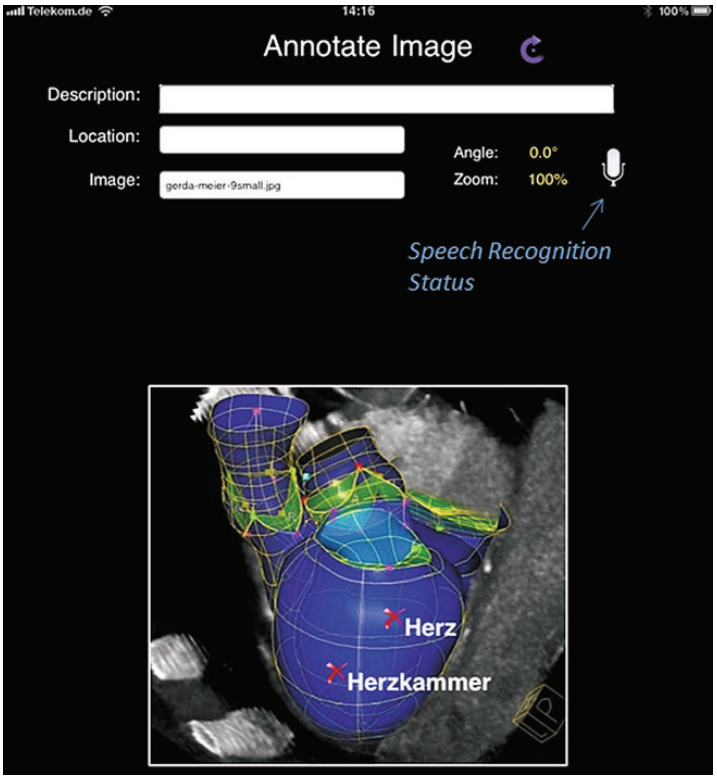

Fig. 3 Speech-based multimodal annotation of medical images

annotation of radiology images by use of speech and gestures.

Also, "Let us consider a collaborative task-based example dialogue." A radiologist treats a lymphoma patient. The patient visits the doctor after chemotherapy for a follow-up CT examination. The focus of the speech-based interactions is to store an RDF-based image region annotation in the image database. The doctor takes care of the patient and image study selection, as well as the image annotation, whereas the dialogue and backend systems "take care" of the image display, the annotation process in form of a database insert query, and the speech confirmation of the successful database transaction that updates the database. An example is shown in following sub-dialogue (for simplicity, the cancer annotation is replaced by a simple anatomy annotation).

1. U: "Show me the CTs, last examination, patient XY."
2. S: Shows corresponding patient CT studies as DICOM picture series and MRI images and MRI videos.
3. U: "Annotate this picture with 'Heart' (+ pointing gesture) and 'Heart chamber' (+ pointing gesture)"
4. S: Shows the new annotations on the image and confirms a database update.

Figure 3 shows the screenshot of the annotation screen. Upon touching a region in the white square, the speech recognition system is activated. After recognition, the speech and gesture modalities are fused into a complex annotation using a combination of medical ontologies. For disease annotations for example, the complete Radlex (http://www.radlex.org/) terminology can be used. More complex interactions can be seen in the demo video,[3] where a set of multi-touch gestures to control the image selection and manipulation phase without the usage of distracting screen buttons is shown. For evaluation, we also developed a desktop-based manual annotation tool called RadSem [21]; anatomical structures and diseases can be annotated while using auto-completion combo-boxes with a search-as-you-type functionality. The resulting annotation is accurate but very time-consuming. Interestingly, the speech and gesture interaction could easily be added to this desktop HCI by using a standard graphical user interface technology, thereby turning it into a proper multimodal HCI with embedded AI technology for speech interpretation and dialogue management (www.dfki.de/RadSpeech/).

4 Ways to Achieve Collaborative Multimodality

There are many good examples of user-centred designs and testing principles of HCI where "direct manipulation" is used to achieve the goal of effective user interfaces. In between the fields of HCI and AI, incorporating domain-specific knowledge into the interface to enhance visual perception [24] (cf. cognitive science) is a relatively new and interesting alternative to the direct application of AI methods for HCI. However, from the AI perspective, it can be argued that advances in AI and human-computer interaction offer unprecedented and seemingly endless opportunities. At the same time, we have to point out that several fundamental scientific and technical impediments must be overcome to achieve these promises [18]. As it turns out, the combination of employed AI technology and HCI, especially in the context of effective retrieval of information while using advanced user interfaces such as multimodal dialogue, is still in a stage of infancy. In the context of information retrieval interaction (also cf. [13]), we often work with systems that:

- cannot recognise and develop common goals
- use canned dialogue segments (e.g., "The answer to your query [input x] is [input y]."");
- only use hardwired interaction sequences (no sub-dialogues or clarifications possible);
- do not use inference services (new information cannot be used to infer new knowledge); and
- have very limited adaptation possibilities (e.g., no context adaptation is possible).

With these limitations, HCI systems cannot conduct convincing mixed-initiative dialogues (cf. fourth principle). We can work against these limitations of current multimodal dialogue and HCI technology by exploiting an approach where

[3] http://www.youtube.com/watch?v=uBiN119_wvg.

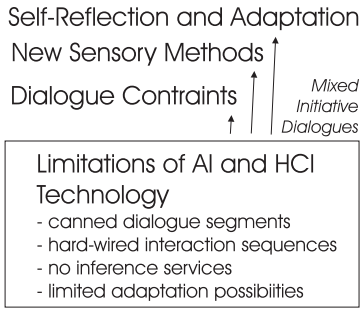

Fig. 4 AI and HCI technology: limitations and challenges

basically three things have to be taken into account which can be thought of as the challenges for the future.

(1) *Obeying Multimodal Dialogue Constraints.* Dialogue constraints are the results of (largely social) rules or norms by which a dialogue participant lives. They play a central role when understanding and creating dialogue acts for a natural dialogue because they account for many of the next user or system moves during the dialogical interaction.

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 conversational maxims, see, e.g., Grice's maxims, and the users' presumptions about utterances; for example, information should be available in an appropriate quantity). Also see [15] for a list of social discourse obligations (Fig. 4).

In the context of multimodal dialogue, extra-linguistic inputs/outputs (i.e., anything in the world outside language and the language modality, but which is relevant to the multimodal utterance) and social obligations have particular multimodal implementations.

Traditional HCI of course turned to the many psychological and empirical works to investigate the acceptance of HCI technology. Extra-linguistic universals in communication and language (also described in the context of politeness, see [2]) are, however, often unrecognised. This is aggravated by the fact that people often cannot predict the behaviour of the HCI, or that the predicted and reached behaviour is not in accord with the social expectations.

In the context of question answering applications and our understanding of intelligent interfaces 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) inform the user about the probability of query success in case of a retrieval task, (3) inform the user as to why the current HCI process is due to fail, (4) balance the user and system initiative. The following lists give examples for important social obligations as dialogue constraints we encounter in terms of natural dialogue when assuming that the conversational goal supports the task of the user (also cf. [30], page 149 ff.). The core social discourse obligations, which can be generated in natural speech or any other modality, are:

1. Self-introduction and salutation: *"Hi there." "Hello." "Good morning/evening." "Hi. How can I help you?" "What can I do for you?" "Hi [name], how are you today?"*
2. Apology: *"I'm so sorry." "I'm sorry." "It seems I've made an error. I apologise."*
3. Gratitude: *"Thank you so much." "I appreciate it." "Thank you."*
4. Stalling and pausing: *"Give me a moment, please." "One minute." "Hang on a second."*

(2) *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 microphone 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.

Attention detection (the technical implementation through on-focus/off-focus) is particularly interesting. 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 the eye-tracker functionality predominantly used in usability studies to learn how to reduce the cognitive load.)

In general, with *anthropocentric* interaction design and models, we seek to build input devices that can be used intuitively. We also recognised that the thumb plays a significant role in modern society—becoming humans' dominant haptic interactor (on mobile phones). This and similar developments should also be reflected in the design of future HCIs; society-based interaction habits might change (e.g., you subconsciously decide to press a doorbell with your thumb instead of the index finger, don't you?). In the context of *anthropocentric* designs, even simple adaptations of well-known interfaces can help a lot. For example, digital pens (see www.anoto.com) and appropriate automatic handwriting and sketch recognition software [9] to provide (new) intuitive input modes (Fig. 5).

(3) *Modelling Self-Reflection and Adaptation.* Humans are able to adapt their dialogue behaviour over time according to their dialogue partners' knowledge, attitude, and competence. This is possible because humans' abilities also include (1) the emotions that are expressed and perceived in

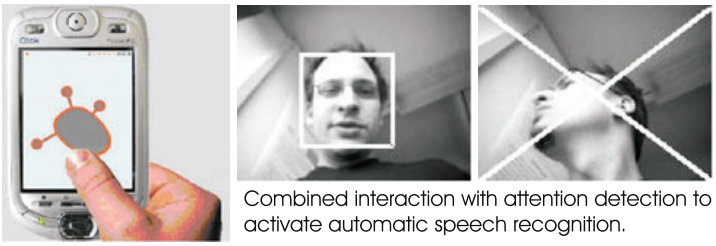

Fig. 5 Anthropocentric thumb sensory input on mobile touchscreen (*left*) and two still images illustrating the function of the On-View/OffView (*right*)

natural human-human communication, (2) the subtle actions and reactions a human dialogue participant performs, and (3) the metacognitive and self-reflective (introspective) abilities of a human dialogue participant to cleverly reason about the actions he or she takes.

Broadly speaking, 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 have had good experiences when using a two-level structure to implement the self-reflection and adaptation mechanism, whereby the cognitive processes are also technically split into two interrelated levels: the meta-level (metacognition) which contains a dynamic model of the object-level (cognition); the two dominant relations between the levels are called control and monitoring [22]. The HCI interaction 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 have been made in the context of dialogue-based question answering. 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 a 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." [29]. Much more research is needed to see if an old idea in human-human conversation will have a new perspective in dialogue-based question answering or other HCI areas with a collaborative goal.

5 Conclusion

We argued in favour of complex AI to be integrated into intelligent user interfaces in order to give HCI researchers an impetus to take a more detailed look into the opportunities of combined efforts. We focused on the concept of collaborative multimodality which states that the user and the system must have a collaborative goal. This allows for a particular view on the HCI areas, centred around the question of how AI might help to solve fundamental problems of HCI technology. Future research should focus more on obeying multimodal dialogue constraints, employing new sensory input and output methods, and re-investigating how to making cognitive architectures available for the HCI community.

These requirements are by no means exhaustive, but do represent a systematic approach to inventing new intelligent user interfaces in the long run. HCI developers could try to tackle the limited opportunities to customise interfaces to specific task and work habits which could be re-used for interface design, ultimately putting AI methods more into the fore and admitting that the trade-off between predictability of a direct-manipulation interface and the convenience of, e.g., predicting the user intent represents a trade-off [16].

Related to users' intentions prediction, a broader spectrum of HCI concerns is conceivable, for example the modelling and evaluating of empathy in embodied companion agents [19], emotion-based reasoning processes in computing for HCI (KI Journal, 25(3)), or complex interaction systems with intuitive capabilities [31]. Clearly, one of the future avenues should rely on affective reasoning which plays an increasingly important role in cognitive accounts of socially intelligent interaction to achieve collaborative multimodality where the computer takes both the form of a "mediator" between human collaborations, and the active role of an authentic collaborator on its own.

"*The question persists and indeed grows whether the computer makes it easier or harder for human beings to know who they really are, to identify their real problems, to respond more fully to beauty, to place adequate value on life, and to make their world safer than it is now*." (*The Poet and the Computer*, Norman Cousins, 1966).

Acknowledgements Thanks go out to Robert Nesselrath, Christian Schulz, Daniel Porta, Markus Löckelt, Matthieu Deru, Simon Bergweiler, Alassane Ndiaye, Norbert Pfleger, Alexander Pfalzgraf, Jan Schehl, Jochen Steigner, Tilman Becker, Gerd Herzog, and Norbert Reithinger for the implementation and evaluation of the dialogue infrastructure. This research has been supported by the THESEUS Programme funded by the German Federal Ministry of Economics and Technology (01MQ07016).

References

1. Anind Dey SR, Veloso M (2009) Using interaction to improve intelligence: How intelligent systems should ask users for input. In: Proceedings of the IJCAI workshop on intelligence and interaction
2. Brown P, Levinson SC (1987) Politeness: some universals in language usage. Studies in Interactional Sociolinguistics. Cambridge University Press, Cambridge

3. Bunt H, Beun RJ (2001) Cooperative multimodal communication. In: Bunt H, Beun RJ (eds) Cooperative multimodal communication, revised papers. Lecture notes in computer science, vol 2155. Springer, Berlin
4. Bunt H, Kipp M, Maybury MT, Wahlster W (2003) Fusion and coordination for multimodal interactive information presentation. In: Stock O, Zancanaro M (eds) Intelligent information presentation. Kluwer Academic, Norwell
5. van Dam H (2003) Dialogue acts in GUIs. PhD thesis, Eindhoven University of Technology
6. Fensel D, Hendler JA, Lieberman H, Wahlster W (eds) (2003) Spinning the semantic web: bringing the world wide web to its full potential. MIT Press, Cambridge
7. Gajos KZ, Wobbrock JO, Weld DS (2008) Improving the performance of motor-impaired users with automatically-generated, ability-based interfaces. In: Proceeding of the twenty-sixth annual SIGCHI conference on human factors in computing systems (CHI '08). ACM, New York, pp 1257–1266
8. Goebel R, Williams MA (2011) The expansion continues: Stitching together the breadth of disciplines impinging on artificial intelligence. Artif Intell 175(5–6):929
9. Hammond T, Davis R (2005) Ladder, a sketching language for user interface developers. Comput Graph 28:518–532
10. Hanson P, Rich C (2010) A non-modal approach to integrating dialogue and action. In: Proceedings of the sixth AAAI conference on artificial intelligence and interactive digital entertainment (AIIDE). AAAI Press, Menlo Park
11. Hitzler P, Krötzsch M, Rudolph S (2009) Foundations of semantic web technologies. Chapman & Hall/CRC, London
12. Horvitz E (1999) Uncertainty, action, and interaction: In pursuit of mixed-initiative computing. IEEE Intell Syst 14:17–20
13. Ingwersen P (1992) Information retrieval interaction. Taylor Graham, London. URL: citeseer.ist.psu.edu/ingwersen92information.html
14. Jameson AD, Spaulding A, Yorke-Smith N (2009) Introduction to the special issue on "Usable AI". AI Mag 3(4):11–16
15. Kaizer S, Bunt H (2006) Multidimensional dialogue management. In: Proceedings of the 7th SigDial workshop on discourse and dialogue, Sydney, Australia
16. Lieberman H (2009) User interface goals, AI opportunities. AI Mag 30(4):16–22
17. Maybury M, Wahlster W (eds) (1998) Intelligent user interfaces. Morgan Kaufmann, San Francisco
18. Maybury MT, Stock O, Wahlster W (2006) Intelligent interactive entertainment grand challenges. In: Proc of IEEE intelligent systems pp. 14–18
19. McQuiggan SW, Lester JC (2007) Modeling and evaluating empathy in embodied companion agents. Int J Hum-Comput Stud 65(4):348–360. doi:10.1016/j.ijhcs.2006.11.015
20. Mitchell TM (1997) Machine learning. McGraw-Hill, New York
21. Möller M, Regel S, Sintek M (2009) Radsem: semantic annotation and retrieval for medical images. In: Proc of the 6th annual European semantic web conference (ESWC2009)
22. Nelson TO, Narens L, Bower GH (eds) (1990) The psychology of learning and motivation: advances in research and theory, vol 26, chap: Metamemory: a theoretical framework and new findings. Academic Press, San Diego, pp 125–169
23. Oviatt S (1999) Ten myths of multimodal interaction. Commun ACM 42(11):74–81 URL: citeseer.nj.nec.com/oviatt99ten.html
24. Paley WB (2009) Interface and mind—a "paper lecture" about a domain-specific design methodology based on contemporary mind science (Benutzerschnittstelle und Verstand – Über eine Design-Methodik basierend auf aktuellen Erkenntnissen der Psychologie). Inf Technol 51(3):131–141
25. Reithinger N, Fedeler D, Kumar A, Lauer C, Pecourt E, Romary L (2005) MIAMM—a multimodal dialogue system using haptics. In: van Kuppevelt J, Dybkjaer L, Bernsen NO (eds) Advances in natural multimodal dialogue systems. Springer, Berlin
26. Reithinger N, Sonntag D (2005) An integration framework for a mobile multimodal dialogue system accessing the semantic web. In: Proceedings of INTERSPEECH, Lisbon, Portugal, pp 841–844
27. Rich C, Sidner CL, Lesh N (2001) Collagen: applying collaborative discourse theory to human-computer interaction. AI Mag 22(4):15–26
28. Russell S, Norvig P (2003) Artificial intelligence: a modern approach, 2nd edn. Prentice-Hall, Englewood Cliffs
29. Sonntag D (2009) Introspection and adaptable model integration for dialogue-based question answering. In: Proceedings of the twenty-first international joint conferences on artificial intelligence (IJCAI).
30. Sonntag D (2010) Ontologies and adaptivity in dialogue for question answering. AKA/IOS Press, Heidelberg
31. Sonntag D (2011) Computing with instinct. Chap.: Intuition as instinctive dialogue. Lecture notes in artificial intelligence, vol 5897. Springer, Heidelberg, pp 82–106
32. Sonntag D, Engel R, Herzog G, Pfalzgraf A, Pfleger N, Romanelli M, Reithinger N (2007) SmartWeb handheld—multimodal interaction with ontological knowledge bases and semantic web services. Lecture notes in computer science, vol 4451. Springer, Berlin, pp 272–295
33. Sonntag D, Reithinger N, Herzog G, Becker T (2010) Proceedings of IWSDS—spoken dialogue systems for ambient environment, Chap.: A discourse and dialogue Infrastructure for industrial dissemination. Lecture notes in artificial intelligence, vol 44. Springer, Berlin, pp 132–143
34. Wahlster W (ed) (2000) VERBMOBIL: foundations of speech-to-speech translation. Springer, Berlin
35. Wahlster W (2003) SmartKom: symmetric multimodality in an adaptive and reusable dialogue shell. In: Krahl R, Günther D (eds) Proceedings of the human computer interaction status conference 2003. DLR, Berlin, pp 47–62

Daniel Sonntag is a senior research scientist at the Intelligent User Interface Department (IUI) at DFKI, lecturer at Saarland University, and associated editor of the German Journal on Artificial Intelligence (KI). He has worked in natural language processing, text mining, multimodal interface design, and dialogue systems for over 14 years. Most recently, he won a German High Tech Award with RadSpeech, a semantic dialogue system for radiologists. His research interests include machine learning methods for multimodal human computer interfaces, mobile interface design, ontologies, and usability.