ERmed – Towards Medical Multimodal Cyber-Physical Environments

Daniel Sonntag

German Research Center for AI (DFKI)
Stuhlsatzenhausweg 3, 66123 Saarbruecken, Germany

Abstract. With new technologies towards medical cyber-physical systems, such as networked head-mounted displays (HMDs) and eye trackers, new interaction opportunities arise for real-time interaction between cyber-physical systems and users. This leads to cyber-physicial environments in which the user has an active role to play inside the cyber-physical system. With our medical application in the context of a cancer screening programme, we are combining active speech based input, passive/active eye tracker user input, and HMD output (all devices are on-body and hands-free) in a convenient way for both the patient and the doctor inside such a medical cyber-physical system. In this paper, we discuss the design and implementation of our resulting Medical Multimodal Cyber-Physical Environment and focus on how situation awareness provided by the environmental sensors effectively leads to an augmented cognition application for the doctor.

1 Introduction

It can be said that most profound technologies are those that disappear by weaving themselves into the fabric of everyday professional life. It would be even better if we could carry and wear those technologies on our bodies which would make us rather independent of the location in which they are used. But what are the requirements for an industry-relevant implementation of new interactive technologies in a cyber-physical system (CPS) in the medical domain?

To answer this question, the following three design aspects have to be taken into account: First, the theory of CPS assumes a complex, safety-critical, intelligent system of interconnected devices. A CPS is a system featuring a tight combination of the system's computational and physical elements. Potential CPS systems include intervention (e.g., collision avoidance); precision (e.g., robotic surgery); operation in dangerous or inaccessible environments (e.g., search and rescue); and last-but-not-least augmented reality, and the augmentation of human capabilities (e.g., healthcare monitoring and decision support for doctors and patients). A medical CPS brings together monitoring devices, such as heart-rate monitors, delivery devices, such as medication infusion pumps, and interaction devices towards an interaction-based multimodal medical cyber-physical environment (CPE). Second, any potential hands-free (multimodal) speech dialogue system solution (such as Smartweb [16] or the now available Siri application [1]) should be extended to include head-mounted displays (HMDs) which

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC 2014, LNAI 8534, pp. 359–370, 2014.

provide new ubiquitous possibilities for interaction-based and real-time cyber-physical systems. Third, which system functionality should be implemented and what is the medical knowledge base in the backend which allows for a seamless, context-based interactive retrieval environment?

In ERmed (`http://www.dfki.de/RadSpeech/ERmed`), we try to give some answers to these questions by designing and implementing a prototypical medical CPE. The main design feature of ERmed is to enter the augmented reality realm, thereby combining multiple input and output modalities (figure 1). Most important are the augmented reality glasses: scrutiny of the eyes of someone engaged in a complex task in a professional (medical) cyber-physical environment should show the potential of vision-based input and output modalities. In this paper, we discuss the design and implementation of a first Medical Multimodal Cyber-Physical Environment and focus on how situation awareness meets mutual knowledge, which is defined by grounded knowledge the cyber-physical environment and the user are both aware of as context features which are obtained from sensor interpretation and database lookups for a resulting medical augmented cognition application.

Fig. 1. ERmed's input and output modalities: RadComet [19], DigitalPen [22], Radspeech [20], ERmed/ERglasses, the focus of this paper according to the emphasis on augmented reality for augmented cognition, MediVa (a touch screen installation), and SmartPen [11]

2 Background

2.1 Multimodal Interaction Systems

The project SmartWeb [16,14] aimed to provide intuitive multimodal access to a rich selection of Web-based information services. The main scenario was a mobile smartphone client interface to the Semantic Web. An advanced ontology-based representation of facts and media structures serves as the central description for rich media content. Underlying content is accessed through conventional web service middleware to connect the ontological knowledge base and an intelligent web service composition module for external web services, which is able to translate between ordinary XML-based data structures and explicit semantic representations for user queries and system responses. The presentation module renders the media content and the results generated from the services and provides a detailed description of the content and its layout to the fusion module. The user is then able to employ multiple modalities, like speech and gestures, to interact with the presented multimedia material in a multimodal way. In project THESEUS [18] we extended SmartWeb's multimodal dialogue system towards integrating the aforementioned semantic web technologies for the Web 3.0 and industrial applications. THESEUS was the German flagship project on the Internet of Services, where the user can delegate complex tasks to dynamically composed semantic web services by utilizing multimodal interaction combining speech and multi-touch input on advanced smartphones. ERmed uses input and output processing components from those projects. The recently launched EU project METALOGUE focuses on natural multimodal interaction. Its goal is to create a multimodal dialogue system where multiple agents expose advanced metacognitive skills that will take into account the user's cognitive model for its exploitation in augmented cognition scenarios. The research project Kognit (BMBF), which uses updated ERmed presentation planning and rendering components, is concerned with fundamental research in augmented-reality based dialogue systems for dementia patients based on augmented cognition and cognitive enhancement.

2.2 Eye Gaze in Intelligent User Interfaces

Motivated by previous findings showing the relevance of eye-gaze in multimodal conversational interfaces [12] and medical application scenarios [2,4] we extended the passive input idea to active user input in the augmented reality realm. This also extends the work of using the gaze information to resolve the ambiguities of users speech [25]. In [5], HMDs have been used in various forms to assist surgeons and other medical personnel to support and improve the visualisation of the workplace related procedures. Commercial see-through HMDs have only recently become available; interestingly, mainstream HMDs (such as Project Glass from Google) do not yet contain eye-tracking hardware. However, this functionality is pertinent in our scenario and the choice of the HMD device and the eye-tracker is very important because of the calibration need [24].

2.3 Focus of Attention Detection and Guidance

Detecting the user's current focus of attention by, e.g., following his or her eye gaze or by interpreting context-based speech commands, allows the system to track the user's activity and intentions, and to verify the system's belief state against the user's mental model. The corresponding concept on the output side—attention guidance—describes the active task of shifting the user's focus of attention towards a particular entity (target highlighting) or away from distracting entities (distraction avoidance). In Ermed, this has been implemented as a prototype with a *mixed reality interface* which uses a synchronised eye tracker and HMD setup [24]; 3D gaze recovery for semantic analysis provides an interesting extension [10].

According to the dialogue environment, the (re)actions to take may be either low-level (e.g., adapt the speech recognition or interpretation confidence threshold to accept a user query) or high-level (e.g., to initiate a system turn or a clarification action, or to maintain mental states according to the underlying dialogue theory and sensory input, such as shifting the focus of attention, indicated by the eye tracker). Other dialogue phenomena such as deixis, turn-taking, and automatic feedback can be modeled with great accuracy. As the gaze carries information about the focus of a person's attention, not only navigation awareness, but even deeper cognitive processes of the user may become visible and interpretable for an AI-based conversational interface or navigation machine. In some way, our head-mounted design with eye-gaze based object/attention recognition paves the way towards machines that "look through the eyes of the beholder." [21]

3 Design of the Multimodal Environment

The following discussion aims to shed light on the aforementioned three design aspects towards the goal of implementing Medical Multimodal Cyber-Physical Environments (CPE), where, in addition to CPS, the environmental aspect (fourth aspect) is key to grounding multimodal interaction and cyber-physical systems in industrial application domains such as medicine: how can experts still control such interactive intelligent environments?

The physical interaction of the users with (interaction) artifacts in the environment must be taken into account. Situational factors of understanding of what the backend retrieval and CPS application engine needs as input turn out to be the pivotal query and decision context factors (context data). From an industrial medical perspective, AI-based proactive HCIs in the form of intelligent user interfaces (IUIs) have not yet been used more broadly because negative side-effects such as diminished predictability of what the system is doing at the moment and lost controllability of the internal multimodal dialogue processes (e.g., a question answering process). These occur more often when (1) AI components are involved, and (2) the context factors for multimodal interaction, integrated storage and knowledge discovery, and intelligent information presentation tasks are not clear. This tendency gives rise to new requirements for usability to account for the special demands introduced by the use of AI for autonomous behavior in the CPE. In the rest of this paper, we focus on the design a d implementation of an interaction lifecycle (figure 2) of our prototypical medical cyber-physical environment (figure 3), taking into account context data from sensors and user interaction (of the caregivers).

We identified the following major challenges and opportunities according to the available real-time CPE scenario background data (figure 3):

- Combine active and passive user input devices in the most convenient way for both the patient and the doctor;
- Provide a direct data acquisition control mechanism and real-time data capture (coordination of distributed input/information streams);
- Use state-of-the-art input and output device strategies: natural speech, graphical head-mounted HCIs, pen-based text and gesture recognition, and eye-gesture-based interaction;

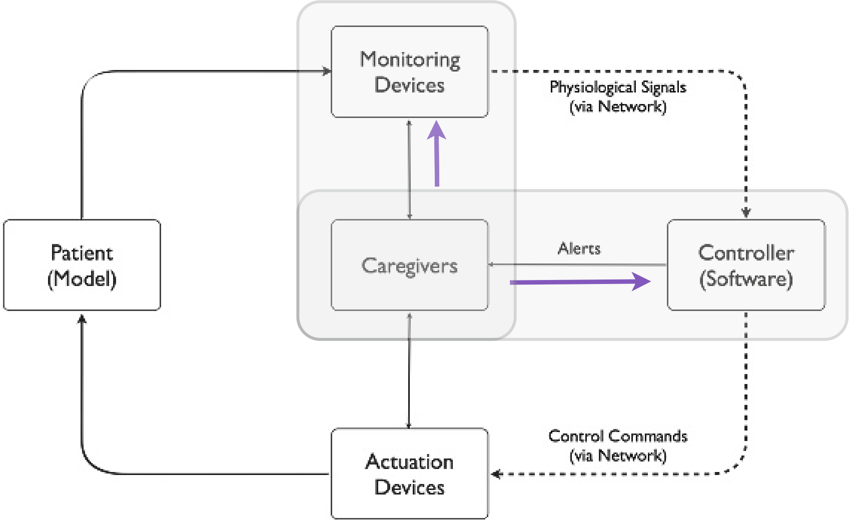

Fig. 2. Medical CPS interaction lifecycle; the purple arrows indicate expert user interaction and intervention

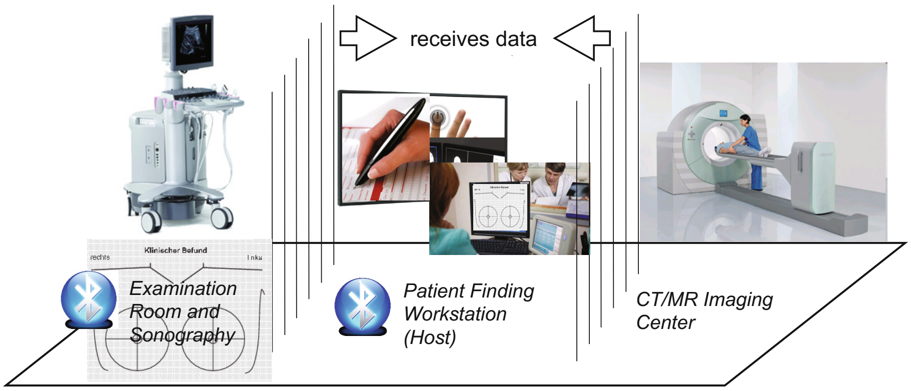
receives data
Klinischer Befund
rechts
links
Examination Room and Sonography
Patient Finding Workstation (Host)
CT/MR Imaging Center

Fig. 3. Real-time CPE background data being received from the medical reporting and examination processes

- Integrate interactive precision-oriented information extraction on textual patient records (text mining);
- Ask the following questions over and over again: can background knowledge describe how a specific medical process is structured—what are the next actions, which information is relevant for the medical expert for a medical decision? Which decisions can the CPE take automatically? What are the possible actions and the "impossible" actions? Which background knowledge provides a set of constraints for the interpretation of the CPE signals?

Within multimodal medical CPE as information systems, we distinguish between data acquisition and data retrieval steps (figure 5). Data acquisition steps aim to capture relevant health data on the basis of a multimodal user interaction, store the captured data in an integrated repository, and extract and semantically label meaningful information units. Due to the sensitiveness of medical data, the data acquisition process is completed by a quality control loop that ensures high quality as well as compliant and consistent data sets. Within the data retrieval step, the existing knowledge repository is accessed to retrieve context-relevant information (for more information, see [23]).

By means of the dedicated multimodal user interaction dialogue we designed, significant context data, such as the name of the patient can be identified. By transforming the extracted context information into a query or filtering request, context-relevant results can be accessed and presented in an intelligent, context-dependent manner in the optical see-through HMD (see Focus of Attention Detection and Guidance).

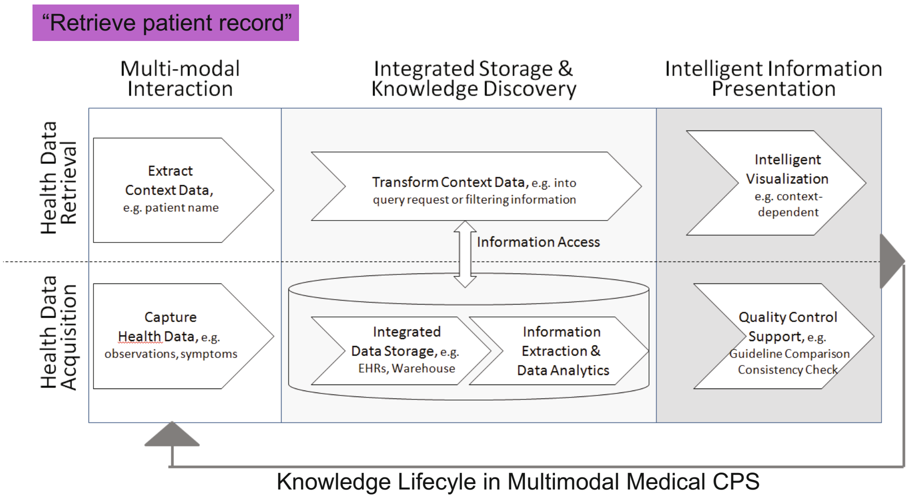
"Retrieve patient record"
Multi-modal Interaction
Integrated Storage & Knowledge Discovery
Intelligent Information Presentation
Health Data Retrieval
Health Data Acquisition
Extract Context Data, e.g. patient name
Transform Context Data, e.g. into query request or filtering information
Intelligent Visualization e.g. context-dependent
Information Access
Capture Health Data, e.g. observations, symptoms
Integrated Data Storage, e.g. EHRs, Warehouse
Information Extraction & Data Analytics
Quality Control Support, e.g. Guideline Comparison Consistency Check
Knowledge Lifecyle in Multimodal Medical CPS

Fig. 4. Knowledge Lifecyle in Medical CPS, image: courtesy of Siemens AG

We conclude the design section by emphasising this binocular view on Medical Multimodal Cyber-Physical Environments: intelligent processing yields the greatest benefits for a combination of interaction and backend processing. It provides avenues for future research in multimodal CPE inside the medical domain context.

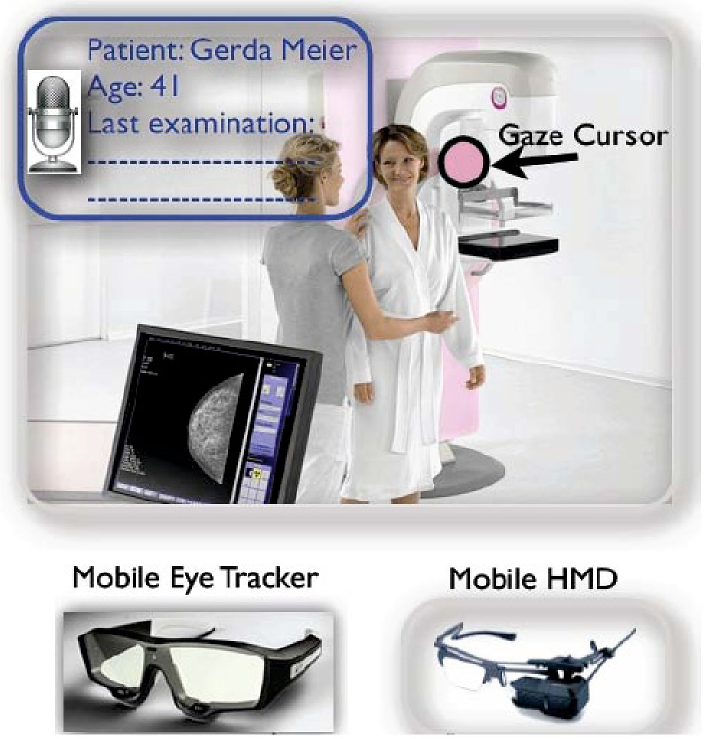

Fig. 5. Augmented cognition application, image: courtesy of Siemens AG [23]

4 Implementation of the Multmodal Medical CPE

For manual data acquisition and documentation purposes in the ”myths of the paperless office [13] digital pens, for example, have been invented and used to implement paper-based interactions in digital environments [17,22]. This is an important step, but the main problem remains, namely that you cannot easily replace a screen-based (laptop) computer because the display is often missing for a proper interaction or retrieval, and time-based transient interaction modes such as speech dialogue cannot replace them properly. ERmed combines multiple on-body input and output devices: a speech-based dialogue system, an HMD, and a head-mounted eye-tracker. The interaction devices have been selected to augment his or her cognition and improve the expert work for a particular application domain, i.e., the physical examination of patients during cancer screening.

4.1 New Multimodal Interaction Possibilities

For the industrial application tasks, we rely on a multimodal discourse and dialogue infrastructure [18] and implemented smartphone gestures interactions in combination with speech (further explained in [20]). New AR interaction possibilities are integrated as follows: essentially, the dialogue context and previous utterances define what the user sees in the mobile HMD. Upon recognition of an object (or person) by the eye-tracker-based image recognisers, we trigger the context-dependent display in the HMD. In addition, we provide new multimodal dialogue interactions by integrating an active learning part into the HMD and eye-tracker scenario:

1. The user gazes at the microphone button and starts the ASR (also cf. [23]).
2. The user says: "learn a new person," which issues a recording command in the multimodal interface and the eye-tracker connection for image positioning.
3. Upon face learning, the dialogue system gets informed about a *new* face and remembers the database instance which is stored in the service backend (takes about 3 sec).
4. The user looks again on the microphone and starts the ASR.
5. The user says: "This (+ active gaze) is a new patient, Peter Meier," which the dialogue system fuses into a face image database command now containing the face classification features and the name of the newly created patient database instance.

Later, the patients can be recognised by both passive or active gaze contingency, i.e., passively or actively (and consciously) controlling a gaze-sensitive button in order to augment cognition by situation-awareness about the participants. This works as follows: if actively looked at a face (dwelled upon for more than 600ms, the name and additional patient info (adapted from Radspeech, `http://www.youtube.com/watch?v=uBiN119_wvg`) is displayed automatically. Currently, we are experimenting with shorter activation times that can be used to "passively" display patient names only. The resulting interaction should, however, be very natural to the user, like looking on thinks and persons and imagining their relevance to the working context. The adapted gaze and speech input provides avenues to make daily routine a bit more effective and yield higher performance outcomes on medical examination tasks or similar knowledge intensive tasks which may benefit from augmented cognition.

In a second implementation step, we have adopted a design direction that allows the activation of the dialogue system while using different types of interaction devices. As a result, the user is able to choose the modality which is most convenient in a specific situation; the following dialogue demonstrates a real-world example (test video available at `http://www.dfki.de/RadSpeech/ERmed`): while the radiologist is analysing medical images application, he or she is requesting for additional information about the patient:

1. The doctor engages the microphone using either eye gaze or pen gestures.
2. Doctor says: "Show me the previous findings in the HMD."
3. HMD: The sight of the doctor is augmented with the corresponding patient file.
4. Text-to-Speech Synthesis: Previous findings: ...
5. The doctor continues with the form-filling process using a digital pen.

6. Doctor uses pen: The medical terms round, smooth, and homogeneous are marked.
7. Text-to-Speech Synthesis: The annotations *round*, *smooth*, *homogeneous* have been recognised
8. HMD: the annotations *round*, *smooth*, and *homogeneous* are highlighted.

5 Research Challenges

5.1 Activity Recognition in a Multiparty Setting

In general, eye tracking technology has been used to help identify persons, for example to evaluate the role of eye-gaze in multimodal reference resolution or provide direct background information about the patients (figure 5). However, the task of conversing with the patient based on eye-gaze patterns introduced a nice idea: the possibility to sense user interests based on eye-gaze patterns and manage computer information output accordingly. Our current approach does not yet differentiate mainly in how the eye-gaze is being recorded and interpreted. In our case, we use a mobile, head-mounted system where the objects are interpreted for environmental communication of the CPE with a focus on face recognition. Further object and activity recognition in a multiparty setting towards collaborative multimodality [15] are still to be implemented.

Remarkable progress in eye-tracking technologies opened the way to design novel attention-based intelligent user interfaces, and highlighted the importance of better understanding of eye-gaze in human-computer interaction and human-human communication within a CPE: a focus of attention of a user is useful to interpret the user intentions, their understanding of the conversation, and their attitude towards the conversation. In human face-to-face communication, eye gaze plays an important role in floor management, grounding, and engagement in conversation [8]. Against this rich background, activity recognition in a multiparty setting could become a break-through technology in CPE to better understand group behaviour and the needs and tasks of multiple participants in a collaborative environment that can include robotic participators and actors as well. For example, two assembly workers work together with a robot arm in a CPE to conduct a highly complex part assembling task: to collaboratively fit together the parts or pieces of complex and heavy compounds. Activity recognition in a multiparty setting supports physical co-operation in the same way as multiparty interaction, turn-taking and conversational roles issues in CPE dialogue management. The envisioned CPE multimodal dialogue extensions include a multiparty setting extension of shared gaze in situated referential grounding [7] and how eye-gaze feedback changes multiparty joint attention thereby extending and generalizing the work of [3].

5.2 Input Fusion, Situation Awareness and Reasoning

Automatically recognised objects are context factors for decision making and physical actions that provide environmental visual cues for activity recognition and augmented cognition. Higher level activities (such as maintaining home appliances and cooking activities) demand for the modelling of domain-specific common sense knowledge (to explain the sub-domain in sufficient detail and provide input for employed reasoning mechanisms) for a better understanding and classification of the result of focus-of-attention detection and interpretation of the eye tracking input (active and passive input). In this regard, the interplay with distributed fusion algorithms and tasks for the highly distributed dialogue CPE will be a central task. Late fusion should allow us to synchronise time-delayed input signals in the highly distributed CPE and use it in the medical CPE, such as demonstrated by on-body multi-input indoor localisation for dynamic emergency scenarios: fusion of magnetic tracking and optical character recognition with mixed-reality displays [9].

6 Conclusion and Outlook

We discussed the ERmed project, the design and implementation of a Medical Multimodal CPE and focussed on how situation awareness provided by the environmental sensors, i.e., the eye gaze input, effectively leads to an augmented cognition application for the doctors. In this scenario, the key factor is to determine the pattern of fixations during the performance of a well-understood task in a professional setting (patient examination), and to classify the types of actions that the eyes perform. Extension of this principle include a more fine-grained classification of eye gaze patterns for multiple more complex CPE tasks that involve augmented cognition. Future work includes activity recognition in a multiparty setting and input fusion, situation awareness and reasoning while following the eye gaze classification of [6] as monitoring functions for more fine-grained augmentation: *locating* objects used later in the CPE process, *directing* the hand or object in the hand to a new CPE location, *guiding* the approach of one object to another (e.g., medical CPE device and patient), and *checking* the state of some CPE variable which is not available in digital form. The monitoring functions will be implemented to implement multiparty communication support and simultaneous dealing with many independent and mobile users or groups of users with a joint intention within a CPE—instead of a single user in the traditional HCI paradigm.

Acknowledgements. This research has been supported by the THESEUS programme funded by the German Federal Ministry of Economics and Technology (01MQ07016). This work was partially supported by EIT ICT Labs and the Nuance Foundation. It is part of MedicalCPS `http://www.dfki.de/MedicalCPS/` and DFKI's ERmed `http://www.dfki.de/RadSpeech/ERmed` projects on the BMBF agenda for Embedded Intelligence in Complex Cyber-Physical Environments (EINCYM). I would also like to thank the associated researchers and developers: Markus Weber (DFKI), Takumi Toyama (DFKI), Mikhail Blinov (Russian Academy of Arts), Christian Schulz (DFKI), Alexander Prange (DFKI), Kirill Afanasev (DFKI), Tigran Mkrtchyan (DFKI), Jason Orlosky (Osaka University) and the medical consultants Alexander Cavallaro (ISI Erlangen) and Matthias Hammon (ISI Erlangen).

ERmed
Erweiterte Realität in der Medizin

eit European Institute of Innovation & Technology
MEDICAL CPS ENVIRONMENTS

References

1. Bellegarda, J.: Spoken Language Understanding for Natural Interaction: The Siri Experience. In: Mariani, J., Rosset, S., Garnier-Rizet, M., Devillers, L. (eds.) Natural Interaction with Robots, Knowbots and Smartphones, pp. 3–14. Springer, New York (2014)
2. Birkfellner, W., Huber, K., Watzinger, F., Figl, F., Wanschitz, F., Hanel, R., Rafolt, D., Ewers, R., Bergmann, H.: Development of the Varioscope AR. A see-through HMD for computer-aided surgery. In: ISAR, pp. 54–59. IEEE (2000)
3. Guo, J., Feng, G.: How Eye Gaze Feedback Changes Parent-Child Joint Attention in Shared Storybook Reading? In: Nakano, Y.I., Conati, C., Bader, T. (eds.) Eye Gaze in Intelligent User Interfaces, pp. 9–21. Springer, London (2013)
4. Ilie, A., Low, K.-L., Welch, G., Lastra, A., Fuchs, H., Cairns, B.: Combining Head-mounted and Projector-based Displays for Surgical Training. Presence: Teleoper. Virtual Environ. 13(2), 128–145 (2004)
5. Keller, K., State, A., Fuchs, H.: Head mounted displays for medical use. J. Display Technol. 4(4), 468–472 (2008)
6. Land, M.F., Hayhoe, M.: In what ways do eye movements contribute to everyday activities? Vision Research 41(2526), 3559–3565 (2001)
7. Liu, C., Fang, R., Chai, J.: Shared Gaze in Situated Referential Grounding: An Empirical Study. In: Nakano, Y.I., Conati, C., Bader, T. (eds.) Eye Gaze in Intelligent User Interfaces, pp. 23–39. Springer, London (2013)
8. Nakano, Y.I., Conati, C., Bader, T. (eds.): Eye Gaze in Intelligent User Interfaces. Springer, London (2013)
9. Orlosky, J., Toyama, T., Sonntag, D., Sarkany, A., Lorincz, A.: On-body Multi-Input Indoor Localization for Dynamic Emergency Scenarios: Fusion of Magnetic Tracking and Optical Character Recognition with Mixed-Reality Displays. In: Proceedings of Pernem/Percom (2014)
10. Paletta, L., Santner, K., Fritz, G.: An Integrated System for 3D Gaze Recovery and Semantic Analysis of Human Attention. CoRR, abs/1307.7848 (2013)
11. Prange, A., Sonntag, D.: Smartphone pen sketch recognition in breast imaging for instant knowledge acquisition. In: Sketch: Pen and Touch Recognition Workshop in Conjunction with IUI (2014)
12. Prasov, Z., Chai, J.Y.: What's in a gaze?: the role of eye-gaze in reference resolution in multimodal conversational interfaces. In: Proceedings of the 13th International Conference on Intelligent User Interfaces, pp. 20–29. ACM, New York (2008)
13. Sellen, A.J., Harper, R.H.: The Myth of the Paperless Office. MIT Press, Cambridge (2003)
14. Sonntag, D.: Ontologies and Adaptivity in Dialogue for Question Answering. AKA and IOS Press, Heidelberg (2010)

15. Sonntag, D.: Collaborative multimodality. KI - Künstliche Intelligenz 26(2), 161–168 (2012)
16. Sonntag, D., Engel, R., Herzog, G., Pfalzgraf, A., Pfleger, N., Romanelli, M., Reithinger, N.: SmartWeb Handheld—Multimodal Interaction with Ontological Knowledge Bases and Semantic Web Services. In: Huang, T.S., Nijholt, A., Pantic, M., Pentland, A. (eds.) AI for Human Computing, LNCS (LNAI), vol. 4451, pp. 272–295. Springer, Heidelberg (2007)
17. Sonntag, D., Liwicki, M., Weber, M.: Interactive paper for radiology findings. In: Proceedings of the 16th International Conference on Intelligent User Interfaces, IUI 2011, pp. 459–460. ACM, New York (2011)
18. Sonntag, D., Reithinger, N., Herzog, G., Becker, T.: A Discourse and Dialogue Infrastructure for Industrial Dissemination. In: Lee, G.G., Mariani, J., Minker, W., Nakamura, S. (eds.) IWSDS 2010. LNCS (LNAI), vol. 6392, pp. 132–143. Springer, Heidelberg (2010)
19. Sonntag, D., Schulz, C.: A multimodal multi-device discourse and dialogue infrastructure for collaborative decision-making in medicine. In: Mariani, J., Rosset, S., Garnier-Rizet, M., Devillers, L. (eds.) Natural Interaction with Robots, Knowbots and Smartphones, pp. 37–47. Springer, New York (2014)
20. Sonntag, D., Schulz, C., Reuschling, C., Galarraga, L.: Radspeech's mobile dialogue system for radiologists. In: Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces, IUI 2012, pp. 317–318. ACM, New York (2012)
21. Sonntag, D., Toyama, T.: On-Body IE: A Head-Mounted Multimodal Augmented Reality System for Learning and Recalling Faces. In: 9th International Conference on Intelligent Environments (IE), pp. 151–156 (2013)
22. Sonntag, D., Weber, M., Hammon, M., Cavallaro, A.: Integrating digital pens in breast imaging for instant knowledge acquisition. In: Proceedings of the Innovative Applications of Artificial Intelligence Conference, IAAI (2013)
23. Sonntag, D., Zillner, S., Schulz, C., Weber, M., Toyama, T.: Towards medical cyber-physical systems: Multimodal augmented reality for doctors and knowledge discovery about patients. In: Marcus, A. (ed.) DUXU/HCII 2013, Part III. LNCS, vol. 8014, pp. 401–410. Springer, Heidelberg (2013)
24. Toyama, T., Sonntag, D., Dengel, A., Matsuda, T., Iwamura, M., Kise, K.: A Mixed Reality Head-Mounted Text Translation System Using Eye Gaze Input. In: Proceedings of the International Conference on Intelligent User Interfaces, IUI 2014. ACM, New York (2014)
25. Zhang, Q., Imamiya, A., Go, K., Mao, X.: Overriding errors in a speech and gaze multimodal architecture. In: Proceedings of the 9th International Conference on Intelligent User Interfaces, IUI 2004, pp. 346–348. ACM, New York (2004)