Towards Medical Cyber-Physical Systems: Multimodal Augmented Reality for Doctors and Knowledge Discovery about Patients

Daniel Sonntag[1], Sonja Zillner[2], Christian Schulz[1],
Markus Weber[1], and Takumi Toyama[1]

[1] German Research Center for AI (DFKI)
Stuhlsatzenhausweg 3, 66123 Saarbruecken, Germany
[2] Siemens AG, Corporate Technology, Otto-Hahn-Ring 6, 81739 Munich, Germany

Abstract. In the medical domain, which becomes more and more digital, every improvement in efficiency and effectiveness really counts. Doctors must be able to retrieve data easily and provide their input in the most convenient way. With new technologies towards medical cyber-physical systems, such as networked head-mounted displays (HMDs) and eye trackers, new interaction opportunities arise. With our medical demo 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.

1 Introduction

In ubiquitous computing, 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. Recent display systems are available as head-mounted displays (HMDs) which provide new ubiquitous possibilities for interaction and real-time systems often referred to as cyber-physical systems [5]. In this paper, we present the design of our multiple device on-body augmented reality interaction system for doctors while examining cancer patients in the medical routine. The augmented reality system comprises a speech-based dialogue system, a head-mounted augmented reality see-through retina display (HMD), and a head-mounted eye-tracker. The interaction devices have been selected to augment and improve the expert work in a specific medical application context which shows its potential. In the sensitive domain of examining patients in a cancer screening programme we try to combine active and passive user input devices in the most convenient way for both the patient and the doctor. The resulting multimodal AR application has the potential to yield higher performance outcomes and provides a direct data acquisition control mechanism. It effectively leverages the doctor's capabilities of recalling the specific patient context by a virtual, context-based patient-specific "external brain" for the doctor

A. Marcus (Ed.): DUXU/HCII 2013, Part III, LNCS 8014, pp. 401–410, 2013.

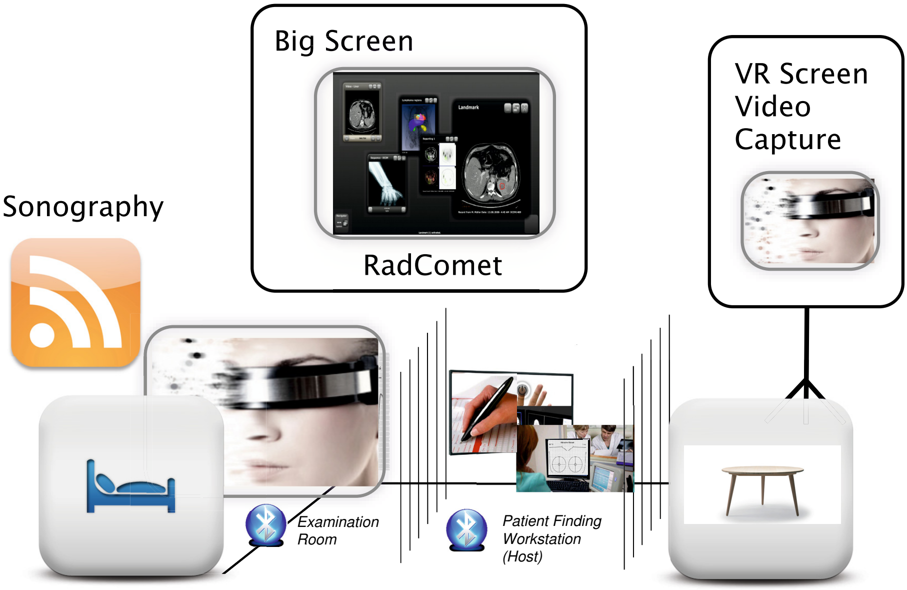

Fig. 1. Medical CPS application environment

which can remember patient faces. In addition, patient data can be displayed on the see-trough HMD—triggered by voice or automatic object/patient recognition. The architecture includes several state-of-the art input and output device strategies: natural speech, graphical head-mounted HCIs, and eye-gesture-based interaction. The prototype of the augmented reality application combines a medical healthcare related (industrial) usecase with multimodal realtime interaction (figure 1). In addition to the doctor's active input at the patient finding station, passive input can be captured from the mobile eye tracker during examination in combination with active speech input. Results can be presented on a big screen (RadComet) or the head mounted augmented reality HMD we will focus on (figure 2). In addition, the picture of the head-mounted camera can be displayed on a large virtual reality (VR) screen for other medical staff members getting access to the real-time video capture. This forms our medical CPS application environment.

2 Background

A cyber-physical system (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

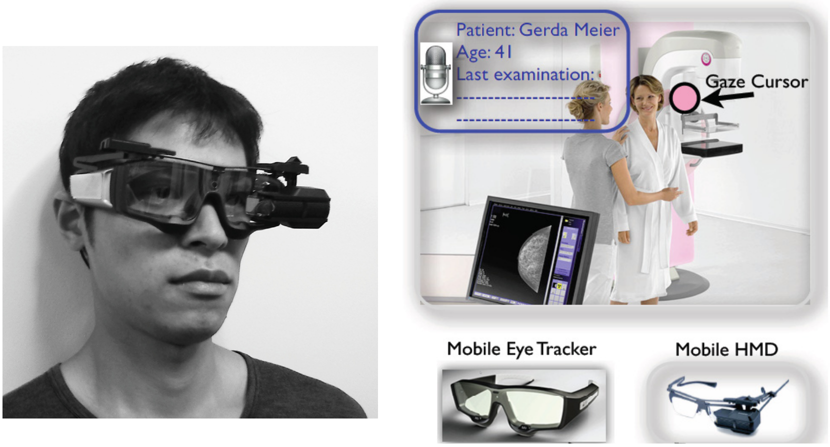

Fig. 2. Mounted eye tracker and HMD combination (left); doctor's view with HMD (right image: courtesy of Siemens AG)

in dangerous or inaccessible environments (e.g., search and rescue); augmented reality, and the augmentation of human capabilities (e.g., healthcare monitoring and decision support for doctors and patients). We bring together augmented reality and augmentation of human capabilities in the mobile medical context: mobile CPS, in which the physical system with a medical purpose has got inherent mobility. This is motivated by the rise in popularity of mobile interaction devices such as smartphones and tablets which has increased interest for CPS developers. As technologies have become small, mobile, and pervasive, the logical next step (beyond the rapid growth of smartphones or tablet PCs or surrounding computers) is the usage of mobile augmented reality and mobile decision support CPS, which will draw people's attention. For example, recent see-through HMD systems (cf. Brother AirScouter or Google Glasses) show massive potential for the future of augmented reality.

The argumentation is as follows: First, mobile eye-tracking glasses allow for mobile gaze-based user input (see figure 2, right: the gaze cursor's position helps to identity the patient's face). Second, semantic sensor data enrichment for HCI has a fundamental role to play (for example, a digital pen can capture the handwriting and interpret it according to a specific patient record [11]). Ideally, this represents the operation of monitoring and interpreting the data coming from mobile sensor in the light of background knowledge expressed in a logical way (ontologies, processes.) Third, background knowledge can describe how a specific medical process is structured, e.g., what are the next actions? What are the possible actions and the "impossible" actions? Which background knowledge provides a set of constraints for the interpretation of the signals for mobile medical CPS?

In [2], the usage of an HMD in ultrasound scanning task has been investigated. [15] enhanced the direct sight of the physician by an HMD overlay of the virtual data onto the doctors view in the context of the patient. In [4], HMDs have been used in various forms to assist surgeons. We provide the first see-through implementation in a multimodal speech-based setting. Over the last several years, the market for speech technology has seen significant developments [7]. In earlier projects [16,8] we integrated different sub-components into multimodal interaction systems. Thereby, hub-and-spoke dialogue frameworks played a major role [9]. We also learned some lessons which we use as guidelines in the development of *semantic* dialogue systems [6]; the whole architecture can be found in [10]. Thereby, the dialogue system acts as the middleware between the clients and the backend services that hide complexity from the user by presenting aggregated ontological data. One of the resulting speech system, RadSpeech [12], is the implementation of a multimodal dialogue system for structured radiology reports. In previous implementation work of a large-scale project[1], we provided a technical solution for the two challenges of speech-based multimodal system engineering and debugging functional modules in domain-specific applications [13]. The next step is the inclusion of a knowledge lifecyle in multimodal CPS.

3 Knowledge Lifecyle in Multimodal Medical CPS

Within multimodal medical CPS as information systems, we distinguish between data acquisition and data retrieval steps (figure 3). Data acquisition steps aim to capture relevant health data on the basis of a multi-modal 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. Again, by means of a dedicated multimodal user interaction dialogue, significant context data, such as the name of the patient can be identified. By transforming the extracted context information into query or filtering request, context-relevant results can be accessed and presented in an intelligent, context-dependent manner (in the HMD).

Described here as two distinguished activities, there is a strong interaction between the two depicted layers. For instance, information extraction can be a acquisition related off-line process. However, within a real-time retrieval application according to a CPS workflow, context data dependent (precision-oriented) information extraction is to be understood as a key element to CPS knowledge discovery. In the following, we will describe the impact and associated opportunities of the health data acquisition and retrieval steps in more detail.

[1] This work is part of THESEUS-RadSpeech (see `www.dfki.de/RadSpeech/`) to implement dialogue applications for medical use case scenarios. It has been supported by the German Federal Ministry of Economics and Technology (01MQ07016).

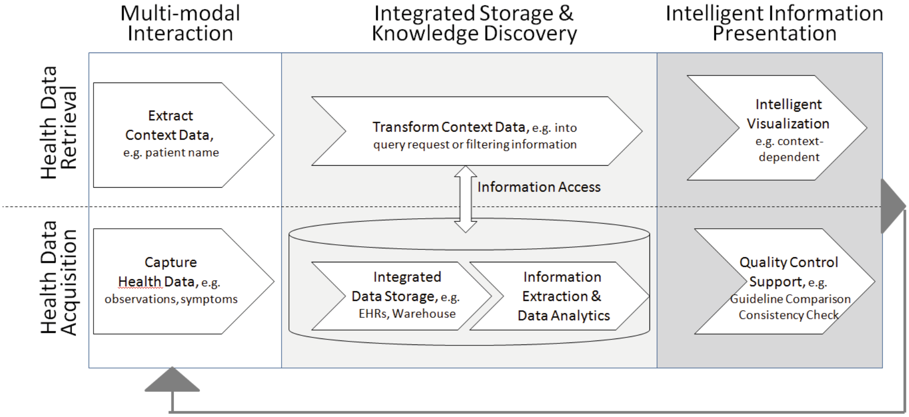
Multi-modal Interaction
Integrated Storage & Knowledge Discovery
Intelligent Information Presentation
Health Data Retrieval
Health Data Acquisition
Extract Context Data, e.g. patient name
Transform Context Data, e.g. into query request or filtering information
Intelligent Visualization e.g. context-dependent
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

Fig. 3. Knowledge Lifecyle in Multimodal Medical CPS

3.1 Health Data Acquisition

The more clinicians know about the health condition of a patient, the better the treatment can be. For instance, comprehensive and high-quality health data of patients can be used to identify patterns within the longitudinal progress of patients health condition; the data can be used to identify silent signals indicating a rare disease or the data can be compared with the health data of related patient populations. However, in todays clinical practice, only a small percentage of patient health data is captured, stored, exchanged, and accessed in a seamless and intelligent manner [14]. The very limited quality as well as density of clinical data is due to several reasons:

1. *Missing IT-Infrastructure*: the overall clinical data acquisition process is often still paper-based and the seamless exchange of data is hindered by either missing or badly implemented interfaces. As of today, a large number of clinical experts share and exchange patient data by scanning and emailing or faxing paper-based documents.
2. *Lack of time*: clinicians do not have "extra time" which they could spend on comprehensive documentation tasks. Usually, the patient interaction—be it an operation or a patient visit—requires almost 100% of their attention. Documentation tasks are only accomplished if they are mandatory, which again is only the case for a very limited number of clinical tasks. The documentation task per se should not draw off the attention of the clinician and this can be achieved along the way without any interruption of the workflow. Speech recognition has been used for that purpose, and we try to extend it to the augmented reality realm.

3. *Complexity of clinical data*: clinical data is very complex. Often, for instance when analysing clinical data with the purpose of conducting retrospective studies, one recognises that particular but important parameters are missing within the collected data sets. In other words, the data collected in clinical routines are usually not complete; particular parameters are not documented as they were not of relevance for a particular case. However, exactly those parameters might be of high relevance when it comes to the comparison of patients with similar diseases or treatments in the context of retrospective studies or analytical applications. Medical CPS architectures should help to elicit the missing information during real-time augmented reality interaction with the doctor.
4. *The value of clinical data increases over time*: Today, longitudinal data is seldom captured for sharing more broadly. The longer the captured time period, the more valuable the data set becomes. However, the majority of todays implemented health systems collect patient data only in the context of a particular treatment episode. Although the documentation of longitudinal patient health data is very promising (in terms of seamless data access and analytics), as of today it is only accomplished for rare or severe diseases. In such cases, the data is maintained/stored within dedicated long-term disease registries, the data collection/acquisition processes are mainly accomplished manually (often on the basis of Phd-studies of doctors-in-training) and implemented as parallel tracks to the clinical routine processes. In other words, the return of investment of longitudinal clinical patient acquisition takes a long time to become measurable.

Comprehensive and high quality clinical data is of high value but it takes some effort to acquire. In the following, we will show how multimodal user interaction in CPS establishes the basis for a seamless health data acquisition process, as well as its intelligent processing and semantic visualisation, and demonstrate its potential to improve the overall quality as well as efficiency of health care delivery. In the long run, also on-body passive sensors should help us collect longitudinal data automatically to be re-injected into the automatic, semi-automatic, manual clinical decision process.

3.2 Health Data Retrieval

The medical application environment comprises of a patient finding workstation (to write digital medical patient reports) and a patient examination room (figure 1). When using our multimodal interaction system, the doctor has the ability to retrieve patient-related information about the patient during the examination and is able to input new data in a convenient way (speech activation with eye tracker) during the examination. When the eye of the doctor focusses on a patient face, recorded faces can be recognised. In addition, new patients can be added by a speech command "*Add new patient x*," while the face is being detected. The retrieval architecture includes virtual databases which can also gather information from web repositories.

First, we want to leverage the doctors capabilities to recall patients and inform himself about a specific patient record. For this purpose, we implemented the first online, head-mounted face learning system which uses a mobile eye-tracker. In the most elaborate interface mode, we allow for a real-time interactive face detection. The face detection scenario is as follows: in a cancer screening programme, which takes place every three months, we evaluated the doctor-patient relationship. As a matter of fact, the probability of being examined by the same doctor in the routine checks is about 10% (due to daily routine personal shift in full treatment hospitals in Germany). As a result, the verification of a patient at the beginning of an examination is a welcome feature for the doctors. In addition, it avoids the need for additional active user input, e.g., a voice command about opening a specific patients record, which can be done automatically. Further, the patient needs not to be interrogated about facts which can easily pop-up automatically in the HMD display.

Second, in the interactive experience with the doctor and the patient, the system should improve the performance of the human-computer interaction and the usability in the patient context. Once the patient is recognised, further speech commands such as show the last CT examination, or what was the last finding allow the doctor to display additional image and text based information in the HMD (privately) or on the big screen when patient should be actively involved in the reading process. Finally, additional annotations and remarks to the specific medical case can be added using speech during the examination without neglecting the patient in, for example, a sonography examination. For this purpose, we use the gaze position on the HMD in combination with an automatic speech recognizer (ASR) as part of the multimodal interaction structure. A little gaze to activate the ASR (and natural language understanding component, NLU) makes the daily routine much more effective and yields higher performance outcomes on the knowledge intensive medical examination and reporting tasks, because the doctor can use the speech system during the sonography examination in a robust way (push-to-talk through the activation), and the results a presented in the see-through HMD (figure 4).

The technical architecture, which includes active and passive input modes (a combination of state-of-the-art components), and a setting for online learning algorithms, is currently under development. Currently we use a simple nearest neighbour search method to recognise faces, but this can be extended to an approximate nearest neighbour method such as in [3] to become productive in a clinical environment. The idea of the ”external brain” for doctors and other hospital staff receives a lot of attention according to our discussions with them. In addition, automatic detection of objects and faces provide a situation context which is very interesting from an academic point of view: which information is adequate in specific contexts to be displayed automatically (system-initiative). It provides avenues for future research in multimodal interaction systems and mobile web applications inside and outside the medical domain context. Two medical storyboards for augmented reality-based CPS interaction have been developed (figure 5).

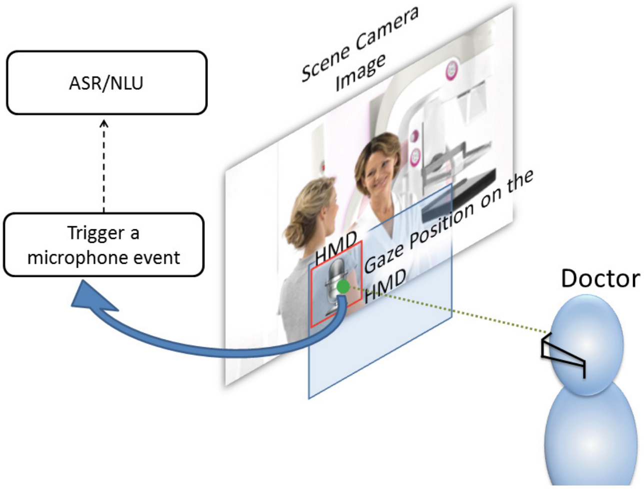

Fig. 4. Patient Examination Scenario where the doctor wears a head-mounted eye-tracker and an HMD

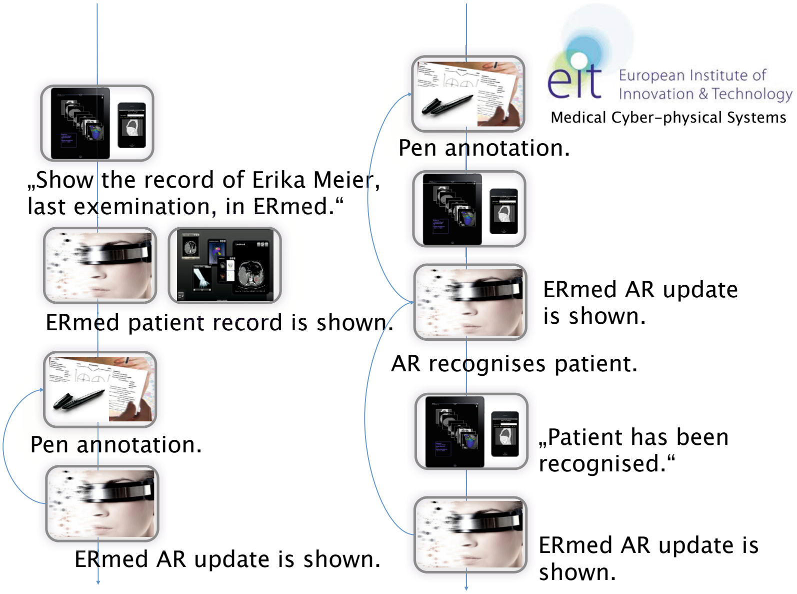
„Show the record of Erika Meier, last exemination, in ERmed.“
ERmed patient record is shown.
Pen annotation.
ERmed AR update is shown.
European Institute of Innovation & Technology
Medical Cyber–physical Systems
Pen annotation.
ERmed AR update is shown.
AR recognises patient.
„Patient has been recognised.“
ERmed AR update is shown.

Fig. 5. Two medical storyboards for augmented reality-based CPS interaction

4 Conclusion

In the RadSpeech use case, we work on the direct industrial dissemination of a medical dialogue system prototype. Recently, structured reporting was introduced in radiology that allows radiologists to use predefined standardised forms for a limited but growing number of specific examinations. However, doctors feel restricted by these standardised forms and fear a decrease in focus on the medical the images [1,17] and the patients. As a result, the acceptance for structured reporting is still low among radiologists for example while referring physicians and hospital administrative staff are generally supportive of structured standardised reporting since it eases the communication with the radiologists and can be used more easily for further processing. In this paper, we extended our first Radspeech scenario (`http://www.youtube.com/watch?v=uBiN119_wvg`) to an augmented reality CPS application. The design of the CPS solution supports the daily routines very well, anticipating todays constraints in hospitals. Doctors are directly put in context and are relieved from manual data capture. It will be interesting to see in future usability studies to what depth doctors will actually apply the sophisticated and facilitated CPS interfaces.

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.

References

1. Hall, F.M.: The radiology report of the future. Radiology 251(2), 313–316 (2009)
2. Havukumpu, J., Vähäkangas, P., Grönroos, E., Häkkinen, J.: Midwives experiences of using hmd in ultrasound scan. In: Mørch, A.I., Morgan, K., Bratteteig, T., Ghosh, G., Svanaes, D. (eds.) NordiCHI, pp. 369–372. ACM (2006)
3. Indyk, P., Motwani, R.: Approximate nearest neighbors: Towards removing the curse of dimensionality. In: Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, pp. 604–613 (1998)
4. Keller, K., State, A., Fuchs, H.: Head mounted displays for medical use. J. Display Technol. 4(4), 468–472 (2008)
5. Lee, E.A.: Cyber physical systems: Design challenges. Technical Report UCB/EECS-2008-8, EECS Department, University of California, Berkeley (January 2008)
6. Oviatt, S.: Ten myths of multimodal interaction. Communications of the ACM 42(11), 74–81 (1999)
7. Pieraccini, R., Huerta, J.: Where do we go from here? Research and commercial spoken dialog systems. In: Proceedings of the 6th SIGDdial Workshop on Discourse and Dialogue, pp. 1–10 (September 2005)
8. Reithinger, N., Fedeler, D., Kumar, A., Lauer, C., Pecourt, E., Romary, L.: MIAMM - A Multimodal Dialogue System Using Haptics. In: van Kuppevelt, J., Dybkjaer, L., Bernsen, N.O. (eds.) Advances in Natural Multimodal Dialogue Systems. Springer (2005)
9. Reithinger, N., Sonntag, D.: An integration framework for a mobile multimodal dialogue system accessing the Semantic Web. In: Proceedings of INTERSPEECH, Lisbon, Portugal, pp. 841–844 (2005)

10. Sonntag, D.: Ontologies and Adaptivity in Dialogue for Question Answering. AKA and IOS Press, Heidelberg (2010)
11. Sonntag, D., Liwicki, M., Weber, M.: Digital pen in mammography patient forms. In: Proceedings of the 13th International Conference on Multimodal Interfaces, pp. 303–306. ACM (November 2011)
12. 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)
13. Sonntag, D., Sonnenberg, G., Nesselrath, R., Herzog, G.: Supporting a rapid dialogue engineering process. In: Proceedings of the First International Workshop On Spoken Dialogue Systems Technology, IWSDS (2009)
14. Sonntag, D., Wennerberg, P., Buitelaar, P., Zillner, S.: Pillars of Ontology Treatment in the Medical Domain. In: Cases on Semantic Interoperability for Information Systems Integration: Practices and Applications, pp. 162–186. Information Science Reference (2010)
15. Traub, J., Sielhorst, T.: Advanced display and visualization concepts for image guided surgery. Display Technology, ... (2008)
16. Wahlster, W.: 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, pp. 47–62. DLR, Berlin (2003)
17. Weiss, D.L., Langlotz, C.: Structured reporting: Patient care enhancement or productivity nightmare? Radiology 249(3), 739–747 (2008)