The Handbook of Multimodal-Multisensor Interfaces: Volume 3

Daniel Sonntag

November 29, 2018

Contents

1 Medical and Health Systems

1.1 Introduction

In this chapter, we discuss the trends of mutlimodal-multisensor interfaces for medical and health systems. We emphasize the theoretical foundations of multimodal interfaces and systems in the healthcare domain. We aim to provide a basis for motivating and accelerating future interfaces for medical and health systems. Therefore, we provide many examples of existing and futuristic systems. For each of these systems, we define a classification into clinical systems and non-clinical systems, as well as sub-classes of multimodal and multisensor interfaces, to help structure the recent work in this emerging research field of medical and health systems.

As discussed throughout this book, multimodal-multisensor interfaces are a major building block in the movement to establish more expressively powerful computer interfaces. In the medical domain, a rapid transformation into this direction is underway, most notably distributed systems that monitor and control multiple aspects of a patient's physiology, health and well-being. In the past, searching and browsing, and discovering patterns in *electronic health records* (EHR) or *electronic medical records* (EMR) and providing operations to align, rank and filter the results, and to visualize data, was the main interest in the human-computer interaction (HCI) community [Wang et al. 2008]. This was complemented by studies of patterns of document and information transfers within the hospital for quality control, see for example Wongsuphasawat et al. [2011]. Today, of particular interest are intelligent user interfaces to reduce healthcare costs in general. These costs are continually increasing, while the available budgets and the number of care-takers are shrinking. For example, in developed countries around the world, an ageing population poses challenges to society, but also unique opportunities for HCI and artificial intelligence (AI) methods in health and wellbeing. To use AI methods to a larger extent, we need a systematic collection of patient information in a digital format. Digital records can be shared across different healthcare settings, store data accurately, and capture the state of a patient across time.

We start by classifying multimodal and multisensory interfaces into clinical systems and non-clinical systems. The clinical view includes, most notably: unobtrusive sensing of vital body signals in clinical environments, data mining of contextual clinical data in different modalities (e.g., clinical records and medical images) and semantic annotation of medical texts and images. Applications include text mining in the health and wellbeing domain, big data analysis and clinical data intelligence, personalized schemes for individualized treatment

and medication, formalizing clinical guidelines for health and wellbeing, as well as decision support. One of the biggest challenges is to reduce the demand for expensive treatments by detecting small physical and mental health issues early. In addition, avoiding larger health problems by clinical treatment or suitable lifestyle interventions. For example, four specific lifestyle factors (not smoking, maintaining a healthy weight, regular exercise, and following a healthy diet) together are associated with as much as an 80 percent reduction in the risk of developing the most common and deadly chronic diseases [Ford et al. 2009]. We can act on this challenge by offering AI-based multimodal-multisensor interfaces for integrating self-monitoring sensors (*quantified-self*). This non-clinical view includes, among other things: smart unobtrusive sensing of vital body signals (of, e.g., care home residents), event and task extraction from life logging by, e.g., video capture, data mining of contextual data, smart coaching algorithms for wellbeing, persuasion technologies, and adaptable interfaces that understand the physical and cognitive abilities of the user.

We focus on clinical and non-clincial systems in sections 1.2 and 1.3, respectively. Three case studies are presented in section 1.4, followed by future directions of multimodal multisensor combinations and virtual reality in section 1.5. For a definition of italicized terms in this chapter, see the Glossary. For other related terms and concepts, also see the textbook chapter on medical cyber-physical systems [Sonntag 2016], the special issue of the KI Journal on health and wellbeing [Gelissen and Sonntag 2015], and the overview of the German flagship project on clinical data intelligence [Sonntag et al. 2015]. Focus questions to aid comprehension are available in this chapter's supplementary digital resources.

Glossary

Medical cyber-physical systems (MCPS) are real-time, networked medical device systems to improve safety and efficiency in healthcare. The specific advantage of the concepts of cyber-physical systems (CPS) involves the use of both real-time sensor devices (e.g., monitoring devices such as bedside monitors) and real-time actuation devices (such as infusion pumps). In this way, MCPS collect information from the monitoring sensors and actuators by, for example, adjusting the setting of actuation devices, firing an alarm, or providing decision support to caregivers. See MedicalCPS [2018] for intelligent user interface projects that fall into this category.

Prevention (primary, secondary, and tertiary) covers several prevention methods: Primary prevention aims to prevent disease or injury before it occurs and includes education about health risk factors. Secondary prevention aims to reduce the impact of a disease or injury that has already occurred and addresses an existing disease prior to the appearance of symptoms. Examples include regular exams and screening tests to detect disease in its earliest stages (e.g., mammograms to detect breast cancer) or diet programs to prevent further heart attacks. Tertiary prevention aims to soften the impact of an ongoing illness

or injury that has lasting effects. Examples are cardiac or stroke rehabilitation programs and chronic disease management programs (e.g., diabetes). New approaches to improve prevention-related user interaction include *persuasive technologies*.

Persuasive technologies focus on the design, development, and evaluation of interactive technologies aimed at changing users' attitudes or behaviors through persuasion, but not through coercion or deception. In general, persuasive technologies are used to change people's behavior. The persuasion approach we support is that choices are not blocked, fenced off, or significantly burdened. The influence on people's behavior in order to make their lives longer, healthier, and better should be subtle. For example, displaying nutrition information at eye-level is a subtle persuasion technology.

Foundational technologies are intruduced in other chapters in Volume 1 and 2 of this Handbook, namely machine learning [Baltrusaitis et al. 2018a, Panagakis et al. 2018], deep learning [Bengio et al. 2018, Keren et al. 2018], and knowledge management [Alpaydin 2018].

Application domains include serious games, conversational agents, or dialogue systems for healthy behavior promotion; intelligent interactive monitoring of patient's environment and needs; intelligent interfaces supporting access to healthcare services; patient-tailored decision support, explanation for informed consent, and retrieval and summarization of on-line healthcare information; risk communication and visualization; tailored access to electronic medical records; tailoring health information for low-literacy, low-numeracy, or under-served audiences; virtual healthcare counselors; and virtual patients for training healthcare professionals.In addition, we address decision support systems especially for the doctor, which model the diagnostic reasoning and decision-making of medical experts, and systems designed to interact directly with patients as healthcare consumers.

An **electronic medical record** (EMR) is a narrower view of a patient's medical history including laboratory values for example, while an **electronic health record** (EHR) is a more comprehensive report of the patient's overall health.

Medical decision support systems are guidance services that predict a patient's health status to influence health choices by clinicians. Other functions can be administrative, but we focus on supporting clinical diagnosis and treatment plan processes by for example proposing medical substances with little adverse effects. Future implementations should be integrated into the clinical workflow, provide decision support such as treatment options at the time and location of care as a MCPS rather than prior to or after the patient encounter, and provide recommendations for care, not just assessments.

Biosignals provide information from a person's biological or physiological structures and their dynamics. Signals measured from the human body typically originate from

neural or muscular activity. Neural activity is captured by methods such as EEG, electroencephalogram, a test that detects electrical activity in your brain using small, flat metal electrodes attached to your scalp. Muscular activity is captured by methods such as EMG, electromyogram, electric signals generated by muscles, or ECG, electrocardiogram, electric signals emitted from the human heart. They are the basis for human computing, physiological computing and affective computing. Also see Silva et al. [2015]. For applications in human computer interaction (HCI) and intelligent user interfaces (IUI), only surface electrodes are used. Signal processing includes, first, time series analysis, and second, the mapping to physical or physiological states [D'Mello et al. 2018, Martin et al. 2018, Schuller 2018, Wagner and André 2018] towards cognitive states [Cohn et al. 2018, Oviatt et al. 2018a, Zhou et al. 2018]. Biosignals of future interest include electric conductance, bioimpedance, and bioacoustic signals.

Telemedicine subsumes physical and psychological diagnosis and treatments at a distance, including telemonitoring of patient functions.

mHealth includes the use of mobile devices in collecting aggregate and patient level health data.

Quantified self is a term used to describe data acquisition on aspects of a person's daily life, e.g., incorporating self-monitoring and self-sensing, which combines wearable *biosignals* sensors and wearable computing.

The **Resource Description Framework** (RDF) is a family of World Wide Web Consortium (W3C) specifications for metadata. It is used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax notations and data serialization formats.

1.2 Clinical Systems

In the future, clinical environments will develop into *medical cyber-physical systems (MCPS)* of their own. This means that patients will get direct treatment according to a direct data acquisition and interpretation workflow. The doctor's decision support will be provided according to the data the MCPS collects from the individual patients. Future MCPS should assist in hospitals, in homes, and other settings [Carayon 2011, Chen et al. 2014a, de Man et al. 2013, Lee et al. 2012].

MCPS involve heavier use of sensors and passive user input in terms of biosignals than traditional multimodal interfaces; hence they do not necessarily require explicit, active input from a user (doctor or patient) with an explicit human-computer interface. The development of multimodal-multisensor interfaces that rely heavily on passive user input processing require *foundational technologies* to be effective and reliable without human control via active user input. Active input modes include speech, hand gestures, eye-tracking, digital pens, smart-

phones and automatic handwriting recognition. Passive input modes include sensors of the clinical environment, biosignals, or smartphone data. Passive input may involve recognition-based technologies (e.g., gesture) or sensor-based information (e.g., acceleration, pressure). This combination of input sources has not yet been explored in medical environments and is of specific interest because it combines previously unconnected modalities and information sources [Sonntag 2016].

A periodic data collection is important for *primary, secondary and tertiary prevention* or monitoring chronic symptoms such as asthma or diabetes. So-called cyber-physical system (CPS) controllers can issue alarms for situations that require attention by a doctor in emergency situations or to let clinicians know about the physiological and emotional state of an individual patient. Figure 1.1 shows the resulting conceptual architecture, including monitoring and actuation devices, a semantic patient model [Sonntag and Porta 2014, Sonntag et al. 2014b], and controller software around the patient and the caregiver. *MCPS* have specific requirements to be met:

1. High confidence medical device software development: This refers to comprehensive verification, validation and testing as well as robustness and fault-tolerance for clinical systems.
2. Anomaly treatment: The modeling of failures, i.e., anomalies, in using and interacting with medical devices, caregivers, and patient behavior must be accounted for.
3. Embedded, real-time, networked system development: This includes architecture, platform, middleware, resource management, QoS (Quality of Service), distributed control and functional programming for future *application domains* of health systems.

One of the major application goals is to issue more accurate and targeted alarms, to let the doctors initiate any necessary treatment immediately. The idea is to bring both patients and caregivers into the controlled perception-action loop around the patient; the controller can also start a treatment autonomously. The main concern besides privacy and security [Friedland and Tschantz 2018] is to avoid false alarms. The long-term direction is to build multimodal-multisensor medical systems that simultaneously sense health status (the state-of-the-art in biosignals processing is covered in volume 2 of this Handbook [Oviatt et al. 2018b]), in order to adapt multimodal communication patterns for user-in-the-loop systems and system responses according to users' status. Although having humans-in-the-loop has its advantage, modeling human behaviors is extremely challenging due to the complex physiological, psychological, and behavioral aspect of human beings [Munir et al. 2013, Wood and Stankovic 2008]. We discuss multimodal and multisensor interfaces separately in order to account for the different needs and challenges in the medical domain.

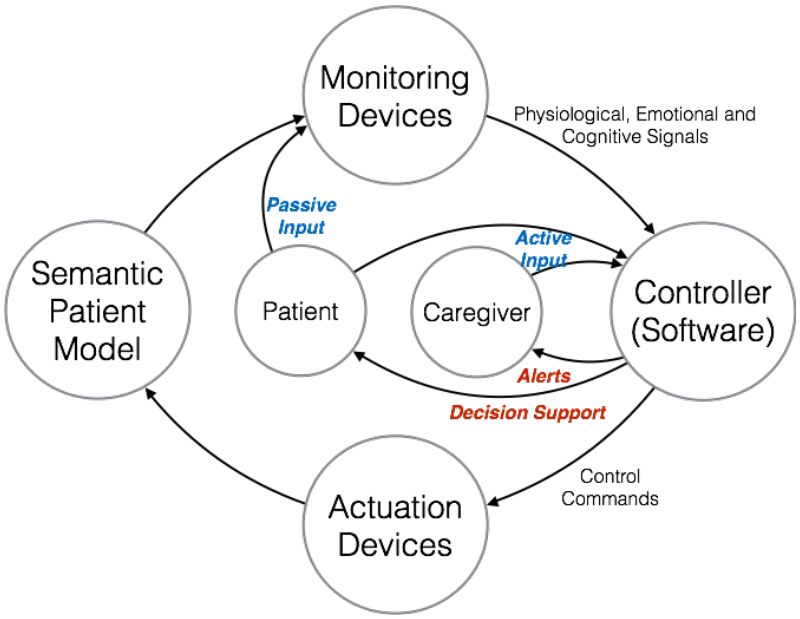

Figure 1.1 Conceptual Architecture: Networked Closed-loop Medical Cyber-Physical System with Human-in-the-Loop Extension. Please note that this extension covers clinical decision support systems for caregivers and patients.

1.2.1 Multimodal Interfaces

Multimodal output involves medical system output from two or more modalities, such as a visual display combined with auditory or haptic feedback, which is provided as feedback to the clinician. More modalities allow for more natural communication, which normally employs multiple channels of expression. It is also the case that more modalities constrain the interpretation and, hence, enhance robustness. For the medical image annotation step for example (see section 1.4.1), predefined speech recognition grammars can be employed. The requirements of medical application domains often include a direct digitalization of multimodal active and passive input data and multimodal feedback in real-time. For example, automatic speech recognition and digital pens allow us to transcribe the clinician's spoken and written input. In addition, the requirements often includes knowledge representation and reasoning about medical concepts. Our design principles can be summarized as follows.

1. Representation and Standards: In a complex medical interaction system, a common ground of terms and structures is absolutely necessary. A shared representation and a common knowledge base ease the dataflow within the system, avoiding costly and error-prone transformation processes [Sonntag et al. 2009]. An ontology-based representation

combines, for example, formal dialogue and image semantics grammars with an *RDF* repository using the SPARQL query standard Sesame [2017]. Linked Open Drug Data (LODD) presents RDF based connected medical information graphs that can serve as knowledge repositories for multimodal user queries. The Life Science Interest Group aims to collect, share, and interlink medical data at very detailed levels by harnessing semantic web technologies [Samwald et al. 2011].

2. Encapsulation: Multiple user interfaces can be connected to the multimodal dialogue system. The system also acts as middleware between the multimodal interface and the RDF repository [Sonntag et al. 2010a].

One implication of recent research findings is that multimodal interfaces are especially well suited for *mHealth* solutions: multimodal interfaces can directly support the multi-functionality of mobile devices and their applications in different application context with different input and output requirements [Oviatt and Cohen 2000]. In mobile settings, multimodal interfaces will promote the multi-functionality of small devices, in part due to the portability and expressive power of multiple input modes. These emerging mobile technologies can be used in extended clinical healthcare (e.g., including blood sugar control, heart frequency, movement pattern after hip surgery) and the computer-mediated communication between doctor and patient.

Multimodal interfaces can also be used in semantic search, as the following examples show: personalized search and summarization over multimedia healthcare information [McKeown et al. 2001]; a multimodal dialogue system for medical images [Sonntag and Möller 2010] which integrates a multimodal interface for speech-based annotation of medical images and dialogue-based image retrieval. In addition, Radspeech [Sonntag et al. 2012] is a speech dialogue example which features mutual disambiguation [Oviatt 1999] of recognition errors. Other systems support multimedia queries in medical search in texts and images, see, e.g., Mourão and Martins [2013]. Luz and Kane [2009] investigate the automatic classification of patient case discussions in multidisciplinary medical team meetings recorded in a real-world setting. More sophisticated passive input methods include accelerating meaningful interface analysis through unobtrusive eye tracking of EMRs towards a quantitative and qualitative assessment of EMR interfaces [Rick et al. 2015]. In a further multimodal interface application, an interactive narrative format for clinical guidelines is presented by Cavazza et al. [2015]. In the future, multimodal input and/or output will have another strong application domain: *telemedicine* should facilitate communication between patients and healthcare providers and doctors. But currently, multimodal interface examples are still missing.

1.2.2 Multisensor Interfaces

Multisensor interfaces in clinical settings are enabling technologies for future telemedicine, integrating biosignal interpretation (covered in volume 2 of this handbook), specialized

doctors at different locations, and robotic surgery. We will focus on robotic surgery after describing clinical applications with sensors for gesture recognition.

Jacob and Wachs [2014] uses sensor-based contextual cues, i.e., gaze, hand position, head orientation, to avoid false positive gesture recognitions for navigating MRI images in the operating room. Jacob et al. [2012] developed a prototype of a robotic surgical nurse for handling surgical instruments in the operating room; experimental results show that 95% of the gestures were recognized correctly. Jacob and Wachs [2014] present a sterile system for navigating MRI images in the operating room. The system has been shown to significantly improve task completion performance. Robotic surgery is another multisensor interface example in the operating room: Along with Taurus [2017] researchers have been developing a dexterous manipulation interface for telepresent surgical robots for remote surgery. This robot is controlled via direct manipulation through two hand controllers which can give visual and tactile feedback. Verbsurgical [2017] is another robotic surgery platform that integrates sensor technologies with medical imaging, data analysis, and machine learning, in order to introduce more autonomy and control of the robotic arms and its end-effectors. In general, robot CPS systems seek to improve minimally invasive and open surgeries (particularly cardio-thoracic, that have so far not benefited from minimally-invasive techniques). The goals are to reduce long hours of operation surgery, increase precision, miniaturize effectors, reduce incision, and decrease blood loss. Additional applications include sensors for example on the effectors for touching soft tissue or bone for computer assisted surgery or even supervised autonomous robotic soft tissue surgery [Shademan et al. 2016]. Current test applications are surgeries in urology, gynecology, general surgery, and thoracic surgery. Turchetti et al. [2012] report on cost evaluation studies of robot-assisted operation. Such operations are compared with those performed by a direct manual laparoscopic approach. Evaluations of the endoscopic procedures using this system suggest that it shortens the length-of-stay in the hospital and reduces recovery times. Critics of the systems targeted at HCI aspects focus on the steep learning curve for surgeons who adopt use of the multisensor environment, in the sense that it is difficult and takes much effort to learn the robot-assisted operations. This means a learning curve with a long, fairly flat region, followed by a big, sudden jump. One gains almost no ability until after 50-70 hours of training. Similar surgical interfaces have been developed: MiroSurge [Tobergte et al. 2011] is a multisensor surgical workstation with several force/torque sensors on haptic input and output devices. It is used in research-based suturing and palpation tasks.

Other multisensor interfaces include robots for basic deliveries or transports of medications, meals, and materials through hospitals, see for example TUG [2017]. In all these multisensor applications it is to be mentioned that crowdsourcing has many options for quality assurance control of medical procedures done with assisted or autonomous robotics, see for example Chen et al. [2014b].

1.3 Non-Clinical Systems

Most non-clinical systems are designed to interact directly with patients. Some non-clinical systems need to understand a patient's intentions, attitude, emotional status, and additional information. In terms of multimodal input processing, facial expression, gaze direction, and emotion as tracked user information are of particular interest. Multimodal input processing helps to provide a holistic view of the patient. Many new medical education systems use multimodal output (e.g., speech and diagnosis graphs). Medical applications are consistent with the general literature on multimodal processing advantages.

Nutrition, physical exercise, and other non-clinical factors contribute to health and well-being. Nowadays, people load their data onto fitness portals such as 'MyFitnessPal,' 'Fitbit', or 'Garmin Connect'. By integrating those resources, some companies have begun to aggregate the data and show analytics with bars or pie charts, thereby providing new opportunities for user-oriented, personalized non-clinical user interfaces. Web-based solutions can feature an incremental knowledge acquisition process with at least two stages, acquiring fitness data with mobile devices and presenting aggregated data in Web portals. As pointed out in Friedland and Tschantz [2018], advances in multimedia content analysis threatens privacy. People must be made aware that non-clinical data collection and analysis can enable unexpected and invasive inferences about people.

Current research and development efforts of non-clinical systems include wearables (digitally enhanced accessories) that are instantiated in familiar real-world objects like watches, wrist bands, digital pens, and tablets. The design question is to make them more useful, versatile, or attractive for digital input processing. As pointed out in Oviatt and Cohen [2015], one advantage of these interfaces is their transparency to users and ability to leverage existing activity patterns, which minimizes a user's cognitive load. One long-term interface design direction will be to combine emerging tangible interfaces that support multimodal input with ones that simultaneously sense users' cognitive load, health status, and similar information in order to adapt system responding to user status. Current non-clinical interfaces for health systems have an emphasis on multisensor interfaces. This extends patient monitoring at hospitals to data collections at home by using portable sensors providing information about a patient's recovery status.

Because EHRs are used to keep track of medications, allergies, conditions, family history, vitals, and exercise, a speech-based question answering system can take this information as additional input. Vitals tracking using smart devices may offer additional sensors. One example is WatsonPaths [Lally et al. 2017], a question answering system that can be asked for the most likely diagnosis or most appropriate treatment, over unstructured information where the answer is not contained in documents. Another application example is GenieMD [2017], a telemedicine platform. One may be interested in side effects of a cortisone shot, or recommendations for available treatment options. Users are able to upload medical records

such as X-rays and lab results for personalized recommendations by answering a short questionnaire relating to chief complaints. The system is not yet multimodal at its interface nor does it use sensor input, but future versions may do so.

To provide a better basis for motivating and accelerating future non-clinical systems for medical and health systems, we provide examples of existing and futuristic systems. All of the medical application examples presented in this section have limitations in scope, but collectively they provide converging perspectives on non-clinical systems towards the design of medical multimodal-multisensor interfaces.

1.3.1 Multimodal Interfaces

We summarize the strengths of multimodal interfaces by providing four examples along four dimensions: multimodal data sources, multimodal interaction, method, and goal.

Sawamoto et al. [2007] explore a method for multimodal interaction logs data sources. Gestures and speech are used. Pattern mining methods are applied to medical interviews in order to extract certain doctor-patient interactions.

Weibel et al. [2013] explore how technology can support natural multimodal interfaces for medical information to provide more effective communication in the medical office. The data sources are EMR interaction logs. The system exploits speech interaction together with sensors to track computer-based activity, visual attention and body movements. The method is pattern mining. The goal is to inform the design of new multimodal healthcare interfaces.

Bickmore et al. [2009] describe an animated, empathic virtual nurse interface for educating and counseling hospital patients in their hospital beds at the time of discharge. It should be emphasized that little research has been done to date on systems to provide information to patients while they are in their hospital beds. Multimodal interaction is provided by a virtual nurse agent (an embodied conversational agent with touchscreen input). The goal is to empower low health literacy hospital patients.

Lisetti et al. [2015] discuss a research project aimed at building socially expressive virtual health agents. Data are collected from interaction logs. They collect data from multiple targets, from obesity to alcohol and drug use, to lack of treatment adherence. Multimodal fusion and fission techniques are primary. The goal is to deliver brief motivational interventions for behavior change in a communication style that individuals and patients not only accept, but also find emotionally supportive and socially appropriate.

1.3.2 Multisensor Interfaces

The focus of this subsection is to summarize the strengths of multisensor interfaces to reduce healthcare costs by results from research and application projects, and to describe how they have been applied to date. In this regard, the present list is by no means exhaustive. Robots and their senso-motoric intelligence are not described [Haddadin et al. 2017]. We focus on activity monitoring of humans by non-intrusive sensors, biofeedback and biomarkers, and

multisensor interfaces in the context of social and virtual companions. In the future, new multisensor applications, especially for diagnostic reasoning, will arise.

1.3.2.1 Activity monitoring of humans by non-intrusive sensors

We focus on sensor-based recognition that can be used remotely from cameras or microphones or sensors that are embedded into devices such as smartphones. Chaurasia et al. [2014] discuss a reminder system for carrying out instrumental activities of daily living (iADLs); the system does not focus on interaction with the user, but instead processes data from a network of sensors. An activity probability model is created to prompt the user via a text interface for the next step in the iADL when inactivity is being observed. Complementary, assistants such as Siri and Google Home (TM) can rely on smartphone sensors. Graus et al. [2016] found out that a smartphone's reminder function is an interesting predictor: the creation time is a strong feature in predicting the notification time, and that including the reminder text further improves prediction accuracy with implications for the design of systems aimed at helping people to complete tasks and to plan future activities. Castro et al. [2015] present a research system to predict daily activities from egocentric images using deep learning. They learn a person's behavioral routines and predict daily activities from first-person photos and contextual metadata such as day of the week and time, or contextual information derived from other sensors. Automatic expressive behavior understanding helps to diagnose, monitor, and treat medical conditions that themselves alter a person's social and affective signals. Valstar [2014] describes automatic behavior understanding, based on multiple sensors. Weiss et al. [2016] compare smartwatch and smartphone-based activity recognition, and smartwatches are shown to be capable of identifying specialized hand-based activities which cannot be effectively recognized using a smartphone. Evaluation results show that smartwatch sensors can identify the "drinking" activity with 93.3% accuracy while smartphone sensors achieve an accuracy of only 77.3%. Maurer et al. [2006] report on medical activity recognition and monitoring using multiple sensors on different body positions for patient monitoring. They focus on sensor fusion. Further examples of user state (e.g., alertness, engagement, physical activity) and trait recognition (e.g., personality, age, gender) where face, fingerprint, and other visual cues are combined, are discussed in Schuller [2018] from Volume 2. Multisensory affect detection is described in D'Mello et al. [2018].

1.3.2.2 Biofeedback and Biomarkers

Integrating biomarkers for mental state detection, for example combining emotional and behavioral indicators for autism detection, represent promising research directions. We describe how multisensor interfaces have been applied to date. New sensor networks including Internet-of-things (IoT) devices may produce new multisensor biomarkers for mental disorders. Typical mental disorders can be detected by new multisensor interfaces in the future. Garbarino et al. [2014] describe a wearable wireless multisensor device for real-time comput-

erized biofeedback and sensor data acquisition. Sriram et al. [2009] propose a mobile medical sensor architecture to provide an efficient, accurate, and economic way to monitor patients' health outside the hospital. They provide arguments that patient authentication is a necessary security requirement in remote health monitoring scenarios. da Silva et al. [2014] present Bitalino, a novel development platform for using biosignals. Their low-cost hardware and open-source software toolkit provides streaming functionality of EMG and EDA (electrodermal activity) to build prototypes for future wearable health tracking devices. Niemann et al. [2018b] use Bitalino to monitor EDA for cognitive assessments for future dementia tests at home. An extended multisensor prototype based on Bitalino will be presented in case study 2 (section 1.4.2). Picard et al. [2017] describe how a commercial wrist sensor reveals sympathetic hyperactivity and hypoventilation in real-time seizure detection by recording wrist motion via 3-axis accelerometer and EDA. Extensions to this work for other mental disorders such as dementia will also be presented in case study 2.

1.3.2.3 Multisensor Interfaces in Social and Virtual Companions

Scherer et al. [2013] describe audiovisual behavior features for depression assessment during multimodal virtual human interviews. They investigated if audiovisual nonverbal behavior descriptors indicative of depression are observable within semi-structured virtual human interview recordings. Additionally, they assessed the correlation of those behaviors with the assessed depression severity. Chen et al. [2016] conduct a study to motivate patients to exercise. They used multisensor fitness trackers including gyroscope, accelorometer, and EDA. Mehlmann et al. [2016] discuss a research project about modeling grounding for interactive social companions based on sensor input, where common ground is needed for joint action and social speech-based dialogue. Further aspects of common ground are planning to achieve joint goals and turn-taking. Especially turn-taking in speech-based dialogue can be informed by sensor input, for example by eye tracking (see chapter 4 in this volume).

1.4 Case Studies

We present three medical case studies in the clinical domain. These are chosen because they describe the transition from monomodal to multimodal applications (first study). The use of sensors helps to interpret a clinical patient's physical state, health status, mental status, and engagement in activities relevant for the assessment and monitoring of pathologies such as Alzheimers (second study). The transition to multimodal-multisensor interfaces is a particularly seminal one in the design of digital tools for behavior characterization in the context of neurodegenerative disorders (third study).

1.4.1 Case Study 1: A Multimodal Dialogue System

In this case study, we present a dialogue system for the annotation and retrieval of medical images where different clinicians are involved in the use of multimodal user interfaces.

1.4.1.1 Background

In contemporary, daily hospital work, clinicians can only manually search for "similar" images using outdated desktop search applications. After considering the relevant categories of similarity, they subsequently apply one filter after the other. For instance, a clinician first sets a filter for the imaging modality (e.g., CT angiography), the second filter for the procedure (e.g., coronary angiography), and so on. In addition to the fact that this approach is quite time-consuming, it is neither possible to formulate complex and semantically integrated search queries in a convenient way, nor can a radiologist easily annotate images with new anatomy or disease information. Hence, the need exists for a seamless integration of medical images and different user applications by direct access to image semantics. Semantic image retrieval should provide the basis for the help in decision support and computer-aided diagnosis.

Our solution is a speech-based dialogue system that integrates a multimodal interface for speech-based annotations of medical images and an image annotation tool for manual semantic annotations on a desktop computer [Sonntag and Möller 2010, Sonntag et al. 2012]. This system implements two of the main applications of medical knowledge acquisition and knowledge integration: first, clinical decision support where some of the clearest opportunities exist to reduce costs by minimising the time for finding treatments based on similar patient cases [Bates et al. 2014], and second, treatment optimization for diseases affecting multiple organ systems. In this case study, we also demonstrate the image retrieval (querying) functionality of the multimodal dialogue interface.

1.4.1.2 Problem Description

Automatic detection of image semantics, i.e., medical annotations of image regions, seems to be feasible, but is too error-prone (at least on the desired annotation level where multiple layers of tissue have to be annotated at different image resolutions or when external expert knowledge is needed). Accordingly, our major challenge is the so-called knowledge acquisition bottleneck. Automatic image recognition cannot easily acquire the necessary medical knowledge about the image contents. As automatic annotation is difficult, we have to address this knowledge acquisition bottleneck problem by concerning ourselves with the question of how to integrate statistical image region annotation (automatic annotation) with manual or semi-automatic annotations.

The requirements discussed with medical experts point to integrate an image annotation tool for annotations on a desktop computer typically performed by medical students (semi-automatic annotation), and a multimodal interface for expert annotations (manual annotation) into a common framework that benefits from manual, semi-automatic, and automatic image annotations. Mainly, a speech-based system for manual annotations of experts should be developed. For new incoming patients, the doctors have to maintain the database and search for similar cases in real-time. Multimodal user interfaces play a significant role in achieving this goal. The system should support the full range of multimodal interaction patterns, such

as deictic or cross-modal references in the context of the annotation process. A remote RDF repository which stores the semantic medical image information and connects the annotation and querying task into a common framework, should make the overall architecture relevant to clinical practice.

1.4.1.3 Solution

For the semantic annotation on a regular desktop workstation, Möller et al. [2009] developed RadSem, a medical semantic annotation and retrieval tool. It consists of a component that implements a method to annotate images, and upload/maintain a remote RDF repository with the images and image semantics. For annotations, RadSem reuses existing reference ontologies and terminologies. More precisely, the Foundational Model of Anatomy (FMA) ontology [Mejino et al. 2008] for anatomical annotations, i.e., annotations of body parts. To express features of the visual manifestation of a particular anatomical entity or disease of the current image, RadSem uses fragments of the RadLex ontology, see Langlotz [2006]. Diseases are formalized using the International Classification of Diseases (ICD-10) [Möller et al. 2010]. Figure 1.2 shows the graphical user interface of the RadSem annotation tool. Images can be segmented into regions of interest (ROI). Each of these regions can be annotated independently with anatomical concepts (e.g., "lymph node"), with information about the visual manifestation of the anatomical concept (e.g., "enlarged"), and with a disease category using ICD-10 classes (e.g., "Nodular lymphoma" or "Lymphoblastic"). However, any combination of anatomical, visual, and disease annotations is allowed, and multiple annotations of the same region are possible. The resulting annotations (mostly anatomical, performed by medical students) are stored in the RDF repository.

In this usage scenario, the expert user—the radiologist—stands in front of the touchscreen installation (figure 1.3, upper part). The interactive system is based on a generic framework for implementing multimodal dialogue systems. Technically, the generic framework follows an object-oriented programming model that eases the interface to external third-party components (i.e., the automatic speech recognizer (ASR) and the text-to-speech synthesis (TTS) component) while using ontology concepts in a model-based design. Several interfaces for the multimodal framework have been implemented: the multimodal touchscreen interface, the event bus, the speech dialogue system, and the application backend as a remote RDF repository. The multimodal touchscreen interface is implemented as a native application using a special window manager for pointing gestures on a touchscreen display. The client provides means to connect to the dialogue system via an event bus, to notify it of occurred events, to record and playback audio streams, and to render the received display data obtained from the dialogue system. The dialogue system contains an ontology-based rule engine for processing dialogue grammars and an external service connector.

The diagnostic analysis of medical images typically concentrates around two questions: i) what is the anatomy? ii) is it normal or abnormal? To satisfy the radiologist's information

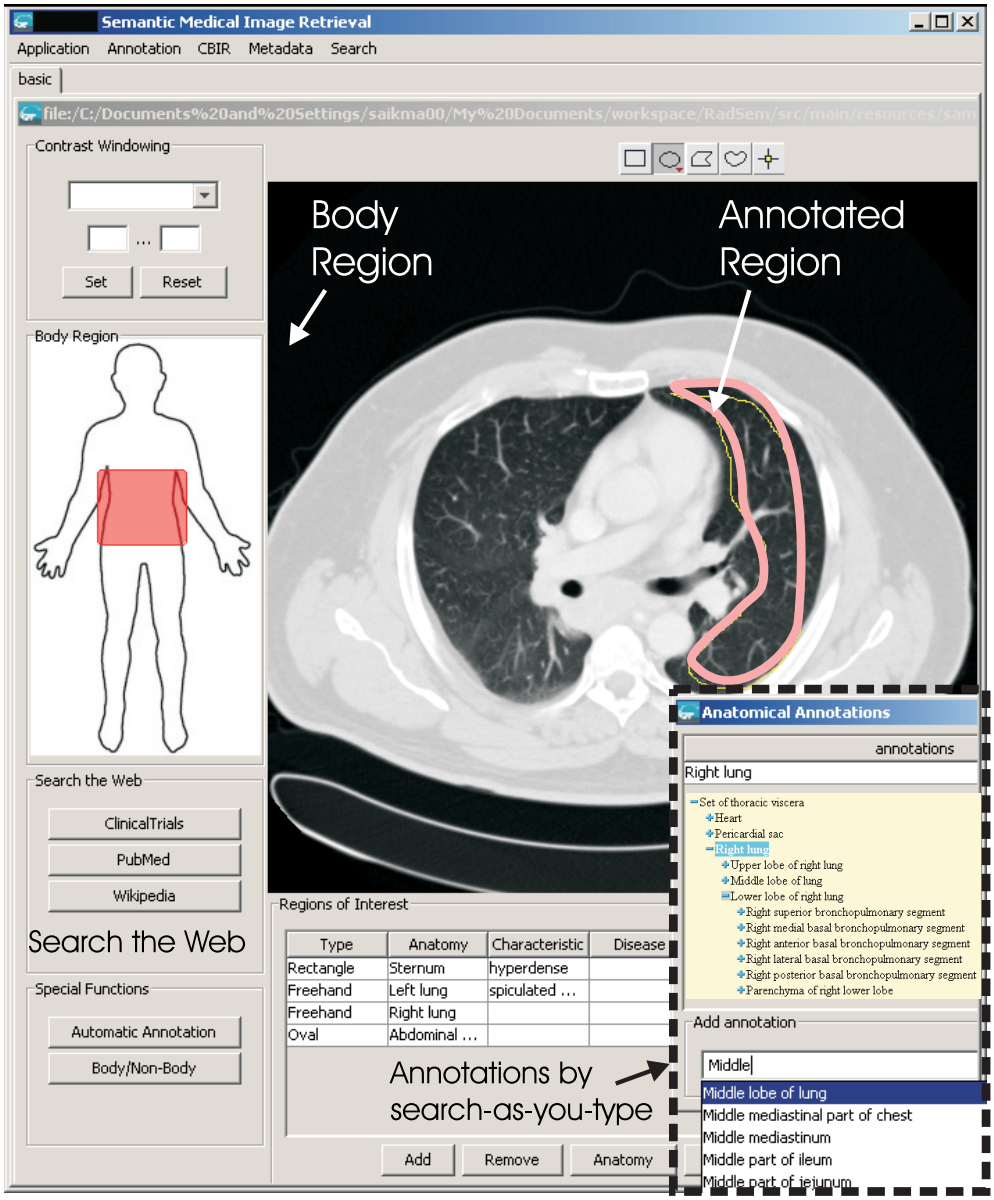

Figure 1.2 Desktop Interface of the Annotation Tool RadSem [Sonntag and Möller 2010] for Manual Semantic Annotations of Medical Images.

need, he or she can formulate the questions in natural speech when a respective image annotation exists. Most importantly, the multimodal interface helps to annotate the respective images and image regions during the patient finding process. Our prototype systems gives a first answer to the following two research questions: first, what kind of information is relevant for the radiologist's daily tasks (a combination of annotation and retrieval). And second, at what stage of the workflow should selected information items be offered and aggregated/annotated in the diagnostic process while using a touchscreen and speech dialogue interface. A multimodal dialogue example is explained in the following:

1. **U:** "Show me the CTs, last examination, patient XY." (retrieval stage)
2. **S:** Shows corresponding patient CT study picture series.
3. **U:** "Show me the internal organs: lungs, liver, then spleen."
4. **S:** Shows patient images according to referral record.
5. **U:** "Annotate this with lymph node enhancement" (+ pointing gesture on region); "so *lymphoblastic*" (expert finding, additional disease annotation (ICD-10)).
6. **S:** "Region has been annotated."
7. **U:** "And replace the characteristic of the other by RadLex: shrunken."
8. **S:** "Region characteristic has been updated."
   → The radiologist switches to another patient (for illustration purposes with a broken finger) and asks for a summary in this additional retrieval stage.
9. **U:** "Give me a summary of this patient." (retrieval stage)
10. **S:** "This is a summary of the fracture: ... "
11. **S:** "Five corresponding CTs will be displayed."
    → The radiologist can now switch again to the differential diagnosis of the suspicious case together with a second medical expert (for the first patient), where the case is examined again and the image annotations can be completed.

A variety of multimodal interaction patterns are implemented in this dialogue, e.g., the resolution of multimodal references. In (5), a deictic reference is resolved (a pointing gesture uniquely singles out an object, it is said to have object-pointing function), whereby in (7) an exophoric reference is given by "the other" annotation already present; it refers to the environment in which the dialogue is taking place, the context of situation is what is displayed on the screen. The command "annotate with" is an implicit reference in the context of the CT image in the current focus. Last but not least, the system builds an own anaphoric reference "corresponding" in (11).

Queries are sent to the open source triple store Sesame [2017]. A direct access to the RDF statements is possible while using the query language SPARQL. This allows us to specify

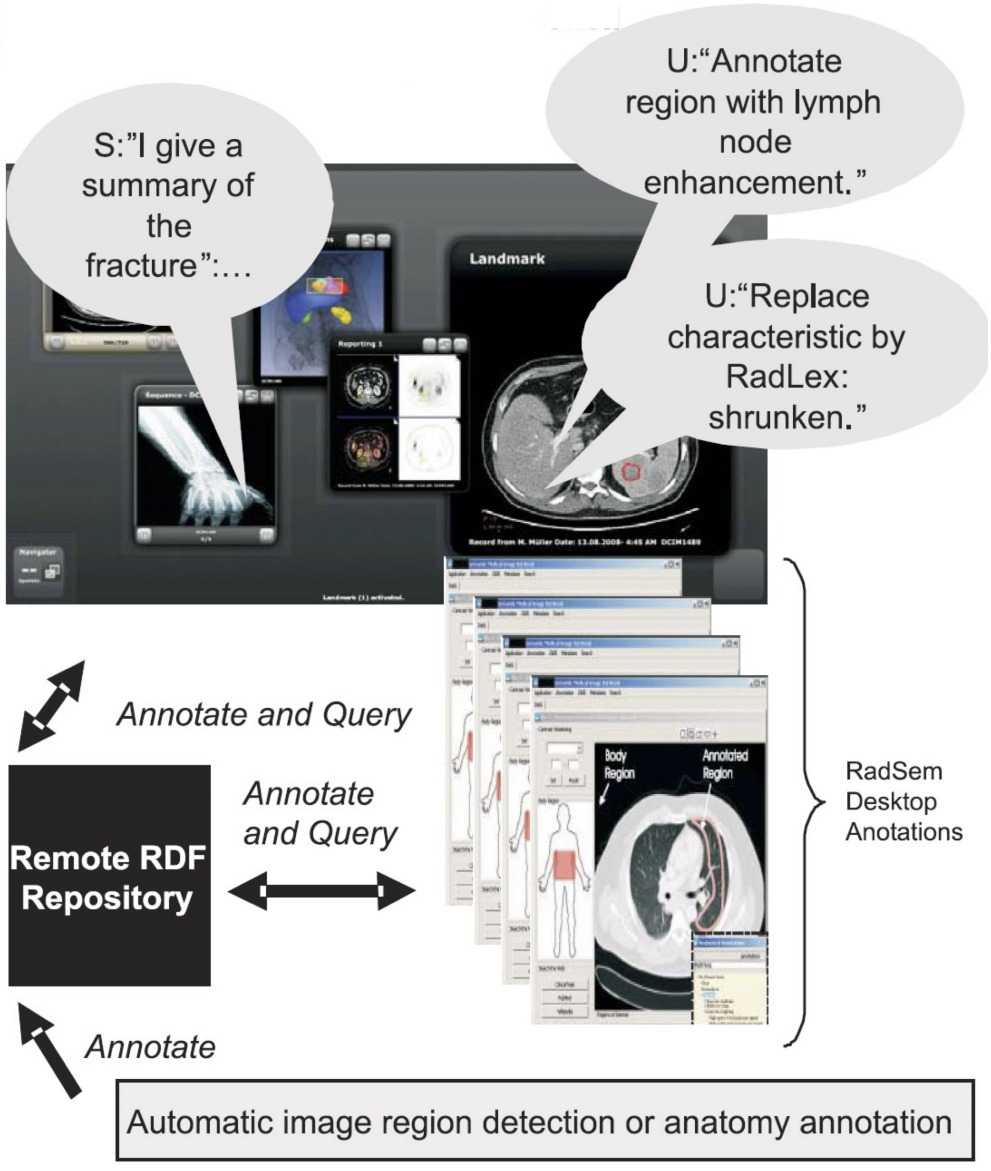

Figure 1.3 Combined Multimodal User Interface for the Semantic Annotation and Retrieval of Medical Images.

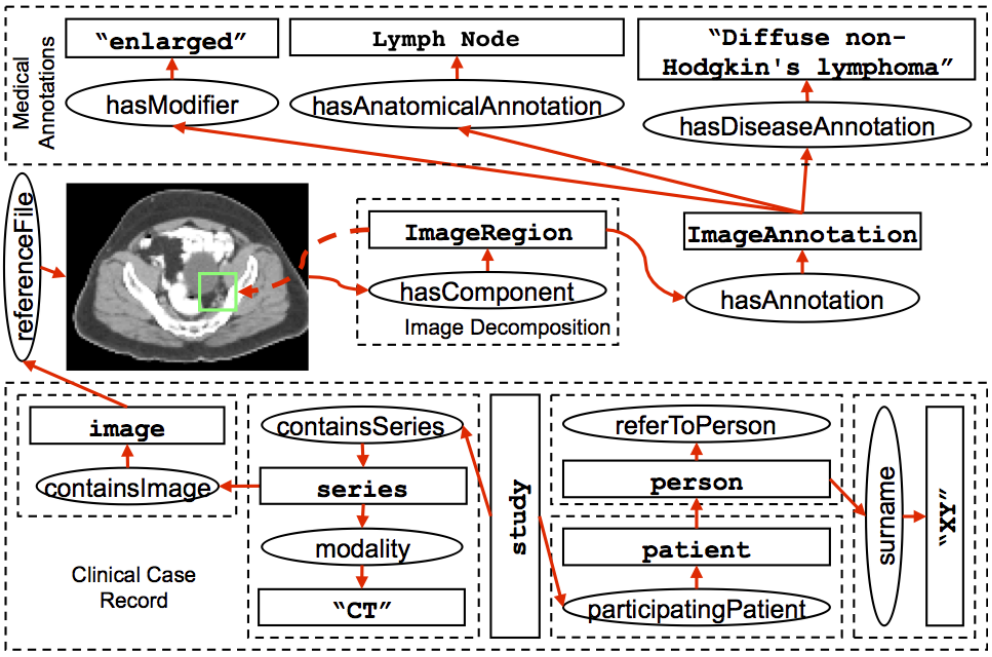

Figure 1.4 RDF Result Graph with Medical Annotations.

queries of almost arbitrary complexity. Queries can span from patient metadata to image annotations to medical domain knowledge and are used to translate the dialogue questions into SPARQL statements. Figure 1.4 shows a graphical representation of the RDF graph that is retrieved when using the query. System responses are based on the retrieved RDF graph. The following SPARQL query example is a translation of the clinician's dialogue question, "Show me the CTs, last examination, patient XY."

```
SELECT ?person ?patient ?imageURL
WHERE {
  ?person mao:surname ?var0 .
  FILTER (regex(?var0, "XY", "i")) .
  ?patient mdo:referToPerson ?person.
[...]
  ?series mdo:modality "CT".
  ?series mdo:containsImage ?image.
  ?image mdo:referenceFile ?imageURL.
[...]
}
```

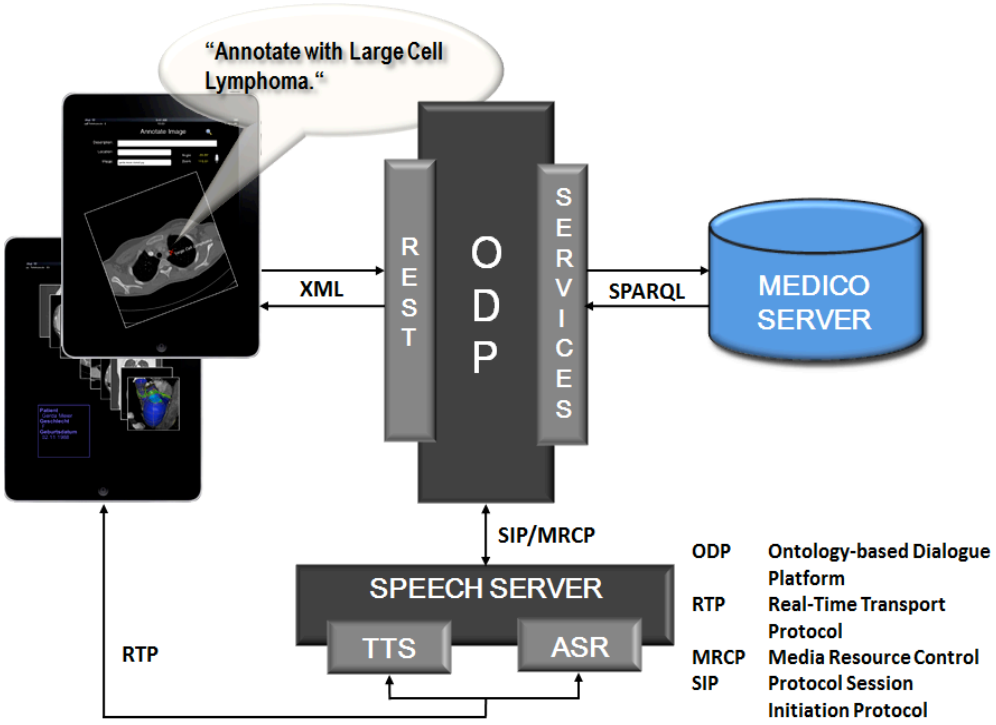

Figure 1.5 Architecture of the Radspeech System showing components and data protocols, see RadSpeech [2011].

While using a tablet for interaction (see the architecture in figure 1.5), additional multi-sensor information from the tablet such as accelerometer or gyroscope information can be used in future radiology interfaces. The communication between the tablet and the dialogue system is based on state-of-the-art web service protocols: the Representational State Transfer (REST) is an established standard that defines a set of constraints to be used for creating web services for distributed information retrieval applications. In summary, the multimodal interface allows the user to annotate medical images with ontology-based medical concepts (RadLex); the annotations are directly transferred to a remote RDF repository. At this point, the radiologist can (1) access the images and image region annotations (a summary can also be synthesized), (2) complete them, and (3) refine existing annotations while using a multimodal dialogue system. Finally, the RDF repository is updated again. Future plans include the inplementation of the speech-based dialogue system in virtual reality where 3D images can be inspected an annotated (also cf. future directions in chapter 1.5).

1.4.2 Case Study 2: A Multisensor Digital Pen Interface

This case study describes a categorisation and implementation of digital pen sensors for behavior characterisation. We focus on the clinical interpretation of time-stamped stroke data from digital dementia tests. Based on using digital pens in breast imaging for instant knowledge acquisition [Sonntag et al. 2014a], where the doctor uses the digital pen for reporting, we now begin to use the digital pen for the patient [Prange et al. 2015].

1.4.2.1 Background

This research is situated within a long-term project Kognit [Sonntag 2015] with the ultimate goal of developing cognitive assistance for patients with automatic assessment, monitoring, and compensation in the clinical and non-clinical context. In the clinical context, we can identify a special target group of interactive cognitive assessment tools as public sector applications: cognitive assistance for doctors in terms of automatically interpreted clinical dementia tests. We think that automatic and semi-automatic clinical assessment systems for dementia have great potential and can improve quality care in healthcare systems. Our new project Interakt [Sonntag 2017] with clinical partners from Charité in Berlin complements previous fundamental research projects for non-clinical interfaces for dementia patients and clinical data intelligence [Sonntag 2015, Sonntag et al. 2015].

Previous approaches of inferring cognitive status from subtle behavior in the context of dementia have been made in a clock drawing test (CDT), a simple pencil and paper test that has proven useful in helping to diagnose cognitive dysfunction such as Alzheimer's disease. This test is the de facto standard in clinical practice as a screening tool to differentiate normal individuals from those with cognitive impairment and has been digitized in a first version with a digital pen only recently [Davis et al. 2014, Souillard-Mandar et al. 2016]. As pointed out in Davis et al. [2014], the use of (1) a digital pen on paper or (2) a tablet and stylus may distort results by its different ergonomics and its novelty. We implement both interfaces for a selection of standard dementia tests in this case study. This should inform future developments of objective neurocognitive testing methods. In particular, we address the issue of what role automation could play in designing multimodal-multisensor interfaces to support precise medical assessments.

1.4.2.2 Problem Description

Neurocognitive testing assesses the performance of mental capabilities, including for example, memory and attention. Most cognitive assessments used in medicine today are paper-pencil based. A doctor, physiotherapist, or psychologist conducts the assessments. These tests are both expensive and time consuming. Further, the results can be biased. In addition to understanding people, their processes, their needs, their contexts, in order to create scenarios in which Artificial Intelligence (AI) technology can be integrated, we are particularly concerned to assess and predict the healthcare status with unintrusive sensors such as those in

digital pens or in tablets. The goal is to improve the diagnostic process of dementia and other forms of cognitive impairments by digitizing and digitalizing standardized cognitive assessments for dementia. Here digitizing is the process of changing from analog assessments to digital forms with hand-writing and gesture recognition. Digitalization is the process to include automatic assessments into the caregiver's task. We aim at weekly procedures in day clinics and base the assessments on clinical test batteries such as the CERAD developed by Morris et al. [1988]. The test is digitized by hand-writing recognition and sketch recognition. Additional new parameters are provided by the digital pen's internal sensors. The conducted cognitive walkthrough for digitalization started with a task analysis with experts at the clinic that specifies the sequence of steps or actions a doctor requires to accomplish a pencil-paper based assessment task as well as the potential system responses to a digitalized version of it. In this case study we identified together with clinical experts that using a digital pen has the following potential benefits:

- the caregiver's time to spend on conducting the test can be reduced;
- the caregiver's time to spend on evaluating the written form can be reduced;
- the caregiver's attention can be shifted from test features while writing (e.g., easy-to-assess completion of input fields) to important verbal test features.
- Digital assessments are potentially more objective than human assessments and can include non-standardised tests and features (for example timing information) whereby previous approaches leave room for different subjective interpretations;
- they can be used to get new features of the pen-based sensor environment, to detect and measure new phenomena by more precise measurement;
- they are relevant for new follow-up checks, they can be conducted and compared in a rigorous and calibrated way;
- they can automatically adapt to intrinsic factors (e.g., sensorimotor deficits) if the user model is taken into account;
- they allow for evidence in the drawing process (e.g., corrections) instead of static drawings that look normal on paper;
- they reduce extrinsic factors (e.g., misinterpreted verbal instructions);
- they can, in the future, be conducted in non-clinical environments and at home.

The challenges we face are three-fold:

1. To identify interface design principles that most effectively support automatic and semi-automatic digital tests for clinical assessments.

2. At the computational level, it is important to investigate approaches to capture both digital pen features and multimodal-multisensor extensions. Some tests assume content features (what is written, language use, perseveration, i.e., the repetition of a particular response such as a word, phrase, or gesture) in usual contexts, as well as para-linguistic features (how is it written, style of writing, pauses, corrections, etc.). These are potential technical difficulties and/or limitations in the interpretation of the results.
3. At the interface level, it is important to devise design principles that can inform the development of innovative multimodal-multisensor interfaces for a variety of patient populations, test contexts, and learning environments.

1.4.2.3 Solution

The scenario includes the doctor and the patient at a table in a day clinic (figure 1.6) which provides most utility. In the following, we focus on the doctor's assessment task. Here, the term utility refers to whether the doctors' intelligent user interface provides the features they need. The conducted cognitive walkthrough started with a task analysis with experts at the clinic that specifies the sequence of steps or actions a doctor requires to accomplish a pencil-paper based assessment task as well as the potential system responses to a digitalized version of it. According to the requirements, we implement a sensor network architecture to observe states of the physical world and provide real-time access to the state data for interpretation. In addition, this context-aware application may need access to a timeline of past events (and world states) in terms of context histories for reasoning purposes while classifying the input data. The result of the real-time assessment of the input stroke data and context data is presented to the doctor in real-time, see figure 1.6. The display includes (1) summative statistics of test performances, (2) real-time test parameters of the clock drawing test and similar sketch tests, and (3) real-time information about pen features such as tremor and in-air time of the digital pen.

Usability design choices, how easy and pleasant the interface is to use, are made according to industrial usability guidelines [Sonntag et al. 2010b] based on usability inspection methods [Nielsen and Mack 1994] and design heuristics based on the psychophysiology of stress [Moraveji and Soesanto 2012]. They can be summarized as follows: For the patient, the digital pen is indistinguishable from a normal pen. So usability is high and (additional) stress is generally low. But the psychophysiology of stress needs to be explored. Lupien et al. [2007] suggest that some of the age-related memory impairments observed in the literature could be partly due to increased stress reactivity in older adults to the environmental context of testing. For the doctor, the psychophysiology of stress needs to be explored, too. There needs to be a possibility to control interruptions (e.g., phone calls) [Moraveji and Soesanto 2012]. In general, for both user interfaces, the effects of stress and stress hormones on human cognition are important. Lupien et al. [2007] enumerate the following stressor characteristics (SC) of

Figure 1.6 Assessment environment with patient and doctor. It also shows the realtime intelligent user interface for the doctor.

interfaces that we use to form further design principles: SC1: Feels unpredictable, uncertain, or unfamiliar in an undesirable manner; SC2: Evokes the perception of losing/lost control. SC3: Has potential to cause harm or loss to one's self or associated objects, living things, or property. SC4: Is perceived as judgment or social evaluative threat including threats to one's identity or self-esteem. Especially SC4 applies in the situation of the patient assessment. Digital pen on normal paper reduces this effect, whereby using a tablet and stylus might increase SC4 stress levels.

The therapist interface, where the real-time interpretations of the stroke data are made available in RDF, is meant to advance existing neuropsychological testing technology according to our interface design principles. Technical details are as follows: First, it provides captured data in real-time (e.g., for a slow-motion playback), and second, it classifies the analysed high-precision information about the filling process, opening up the possibility of detecting and visualising subtle cognitive impairments; also it is zoomable to permit extremely detailed visual examination of the data if needed (as previously exemplified in Davis et al. [2014]).

name	pen input	symbols
AKT - Age-Concentration	100%	cross-out
CDT - Clock Drawing Test	100%	clock, digits, lines
CERAD - Neuropsychological Battery	20%	circles, rectangles, cubes, etc.
DemTect - Dementia Detection	20%	numbers, words
MMSE - Mini-Mental State Examination	9%	pentagrams
MoCA - Montreal Cognitive Assessment	17%	clock, digits, lines
ROCF - Rey-Osterrieth	100%	circles, rectangles, triangles, lines
TMT - Trail Making Test	100%	lines

Table 1.1 Comparison of the most widely used cognitive assessments

Multimodal-multisensor extensions can be implemented with a tablet device (figure 1.7). Additional modalities can help in the analysis of observed user behavior. When interacting with a tablet computer, multiple built-in sensors can be used in addition.

Besides pen-based input, we consider eye tracking and facial expression analysis via the video signal of the front-facing camera, natural speech captured by the built-in microphone, and additional sensor inputs of modern tablet devices. RGB-based eye tracking is interesting for multimodal interaction with a tablet, because it is deployable using the built-in front-facing camera. However, this gaze estimation is erroneous which should be considered in the interaction design [Barz et al. 2018]. OpenFace [Baltrusaitis et al. 2018b] is an open source toolkit for facial behavior analysis using the stream of an RGB-webcam. It provides state-of-the-art performance in facial landmark and head pose tracking, as well as facial action unit recognition which can be used to infer emotions. The openSMILE toolkit [Eyben et al. 2013] provides methods for speech-based behavior analysis and is distributed under an open source license. It offers an API for low-level feature extraction from audio signals and pre-trained classifiers for voice activity detection, speech-segment detection and speech-based emotion recognition in real-time.

The implemented pencil and paper tests are shown in table 1.1, namely AKT [Gatterer et al. 1989], CDT [Freedman et al. 1994], CERAD [Morris et al. 1988], DemTect [Kalbe et al. 2004], MMSE [Folstein et al. 1975], MoCA [Nasreddine et al. 2005], ROFC [Canham et al. 2000], and TMT [Reitan 1992].

The pencil and paper tests have been transferred one-to-one, meaning that the digital versions of pen input fields look just as the analog versions. Table 1.1 shows the absolute percentages of the test questions where the pen is used to answer them. The selection of the tests accounts for a variety of patient populations and test contexts. Concerning the test context, a doctor can always switch between the digital pen and the tablet and stylus

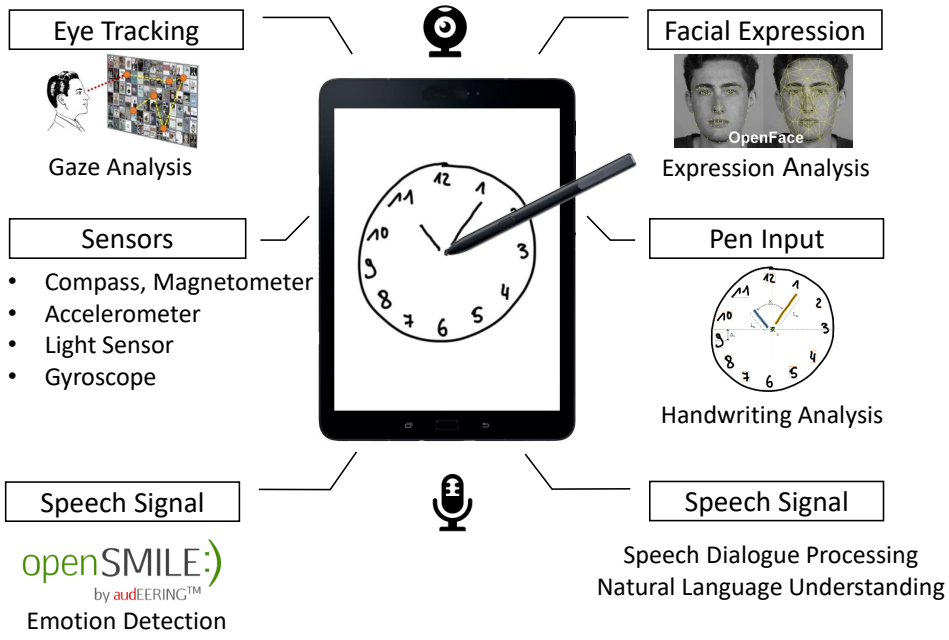

Figure 1.7 Multimodal-multisensor tablet device.

version. The tablet version can always use multimodal-multisensor input to cover additional test contexts.

1.4.2.4 Lessons learned

In this section we discuss which choices we have made in the first 18 months of the project Interakt, the analysis of alternatives considered, as lessons learned. We focus on specific designs and decisions that reduce the potential for failures when considering similar applications.

1. The primary motivation of using a digital pen on normal paper stems from the spatial and temporal precision of the obtained stroke data which provides the basis for an unprecedented degree of precision during analysing these data for small and subtle patterns; classifying the strokes for their meaning is a sketch interpretation task in addition. As a result, we can get assessment data based on what is written or sketched, and how the spatio-temporal pattern looks like. The alternative of using a tablet and stylus turned out to be an additional stress factor for both patients and doctors, as first formative

evaluations suggest. As a result, the formative evaluations with patients will be done on the digital paper version. This choice restricts the possibility to gather multimodal data from a tablet, which provides the same spatial and temporal precision of the obtained stroke data.

2. While the tablet version is not always the first choice, the technical implementation is much easier than the digital pen on normal paper version. The reason is the complicated software development kit (SDK) for creating the digital paper forms on normal paper.
3. How will the data from the experiment be gathered without violating privacy regulations [Friedland and Tschantz 2018]? For example, video capture is currently not allowed. In addition, we need a method to capture assessment results (or corrections/comments) from the doctor while he or she is using the doctor's interface. Will it interfere with the anticipated normal use? Here, the enumeration of the stressor characteristics need to be completed and turned into interface design principles.
4. A version for self-assessment at home for the patient needs to have an ability to control interruptions (e.g., phone calls) [Moraveji and Soesanto 2012].
5. The digitalisation of widely used cognitive assessments has four consecutive steps: first, the one-to-one transfer from a paper and pencil test to a digital version; second, the selection of pen features that are relevant for the classification task; third, the adaptation of the caregivers' instructions to include automatically interpreted test results. And fourth, the inclusion of multimodality and multisensor data for additional test parameters.
6. Digital assessments allow for evidence in the drawing process (e.g., corrections) instead of static drawings that look normal on paper. Doctors need to be instructed when to use the slow-motion playback function. To automatically propose replaying a writing scene for further inspection is another interesting classification task where the system can take initiative.
7. The coverage of implemented tests is rather independent of the availability of suitable patient populations and test subjects. It is rather difficult to get the critical amount of conducted tests for machine learning experiments to find subtle patterns that are sensitive or specific to dementia assessment.

Using digital pens for the assessment of dementia can be generalized in several ways, most notably for use by those in the cognitive impairments field. Digitalized dementia tests can be used for the detection of other neurodegenerative diseases such as Parkinson. Some of the described tests in table 1.1 have already been used in this direction, such as MMSE and a more sensitive similar test MoCA. In addition, this work could help returning veterans suffering from traumatic brain injuries (TBI). J Wagner et al. [2011] used CDT to assess cognition and predict inpatient rehabilitation outcomes among persons with TBI. Doctors working in

inpatient neurorehabilitation settings are often asked to evaluate the cognitive status of persons with TBI and to give opinions on likely rehabilitation outcomes. In this clinical setting, several other digital pen tests could be used for cognitive assessment and outcome predictor among inpatients receiving neurorehabilitation after TBI. It should be possible to better monitor the rehabilitation outcome. As explained above, digital assessments could be relevant for new follow-up checks. They can be conducted and compared in a rigorous and calibrated way.

Future research in the clinical domain includes pen-based assessments to treat patients in an automatic fashion and from multimodal input. Conserning multimodal input, for interpreting verbal utterances of the CERAD test battery for example (therapists have problems in taking notes of user answers and comments while conducting a test), a dialogue framework can be used in the future. Combining active speech and pen input should, in the future, be explored towards multimodal approaches to determining cognitive status. This can be done through the detection and analysis of subtle behaviors and skin conductance sensors.

Using digital pens for the assessment of dementia can be generalized for use by those outside the cognitive impairments field. Current research investigates the use of handwriting signal features to predict domain expertise in several educational contexts [Oviatt et al. 2018c]. The trend towards multimodal learning analytics becomes apparent, where natural communication modalities like writing (or speech) are complemented with gestures, facial expressions, and physical activity patterns. The combination of our low-level stroke features with selected components of the implemented cognitive tests, together with the domain expertise prediction task in Oviatt et al. [2018c] might open up opportunities to design new educational technologies based on individualized writing data resulting in better user modeling.

1.4.3 Case Study 3: A Multimodal-Multisensor Framework

In this case study, we report on a multimodal multisensor framework for recording and analyzing handwriting input that is captured using a digital pen (cf. case study 2) and electrodermal activity (EDA) captured by the Bitalino sensor board for extending behavior characterisation for cognitive assessments in cognitive impairment cases.

1.4.3.1 Background

In the future, large scale community screening programs can arise from multimodal data collections to identify profiles of impairment across different cognitive, psychiatric and functional disabilities. Multimodal-multisensor data guide differential diagnosis and further assessment, because digital assessments are unbiased to a large degree. Stress and emotion changes reflect the activity of the sympathetic branch of the autonomous nervous system [Boucsein 1992]. Because sweat is an electrolyte solution, changes in the sweat level lead to changes in the skin conductance or electrodermal activity. Changes in EDA, especially the skin conductance response (SCR) can be used to detect stress, affect, and arousal [Kurniawan

et al. 2013, Pecchinenda 1996, Saitis and Kalimeri 2016, Zhai and Barreto 2006]. Because of this correlation of stress (and cognitive load) and EDA, we believe it to be a suitable tool for the indication of cognitive impairments, in combination with digital pen features.

1.4.3.2 Problem Description

To include EDA into future digital pen based screening methods is very interesting because it is a process tracing method (unobtrusive and continuous measure) for neural activity and can reflect psychological processes, but context and sensor fusion is needed because it is a multifaceted phenomenon (sensitive but not specific). The digital pen-based environment provides such a sensor fusion context for its interpretation. Towards a multimodal cognitive assessment framework, three challenges need to be addressed. First, the selection of a useful subset of multimodal digital tests together with their implementations. Second, the inclusion of EDA and pen data into a multimodal-multisensor platform. And third, the study design for a future multimodal and digital cognitive assessment framework.

1.4.3.3 Solution

One of the most often used assessments for cognitive impairment is the CDT (also see case study 2), where the subject is asked to draw an analog clock including the numbers of the clock face and a specific time [Freedman et al. 1994]. Most of the existing tools concentrate on automating existing scoring schemes, which were originally designed to be processed and evaluated by therapists. In this described solution, Niemann et al. [2018a] combine EDA with digital pen data for a direct interpretation of writing behavior and biosignals: In addition to the traditional assessment categories (e.g., clock face numbers being in the correct place) they also take into account the EDA sensor data. In order to do so, writing tasks need to be split into semantic categories first. Figure 1.9 shows the visualization of a selected semantic feature set in the context of CDT. Participants are asked to draw a clock face with the time set to 10 past 11 o'clock. The drawn clock is then examined by a trained physician and rated based on a predefined scoring scheme, reflecting the visual appearance and integrity of the clock using a numerical score. In CDT, we extract the following semantic features from the traditional scoring system:

- c denotes the center point of the clock (centroid), the closer it is to the center of the clock's circle, the more points are awarded.
- L_h and L_m represent the lengths of the hour and minute hands respectively. If the clock is well drawn, the hour hand should be shorter than the minute hand.
- The angle between the hour and minute hands is denoted as α, together with the orientation of the hands it can be used to determine if the correct time was set.
- Δ_9 is the displacement of clock face digits relative to their ideal location. In this example it is the vertical offset of digit number 9 to its correct center position.

name	approx. time needed	speech input	pen input	EDA input	evaluation
AKT	15 min	NO	YES	YES	i-scores
DemTect	6-8 min	YES	YES	YES	s-scores
MMSE	5-10 min	YES	YES	YES	s-scores
TMT	3-5 min	NO	YES	YES	s-i-t-scores
CDT	2-3 min	NO	YES	YES	i-t-scores

Table 1.2 Comparison of the most widely used cognitive assessments concerning multimodal input.

With the help of semantic features and appropriate time stamps, the EDA signal can be fused with the digital pen interpretation and high EDA amplitudes traced back to writing and sketching tasks. In other words, the deviation of semantic feature interpretations from the norm indicate stress and cognitive load points that can be synchronized and validated by the sensitive, but non-specific EDA signal. Similar semantic features can be extracted from similar tests [Prange et al. 2018b]. The test overview in table 1.2 includes how much time the assessment usually takes, multimodal-multisensor aspects of active/passive pen input, active speech input, and passive EDA input, as well as how the assessment is evaluated by experts. The selection is based on the trade-off between the amount of tasks including handwriting input, the amount of movement during tasks (which might interfere with EDA signal data), and the need for additional speech input by the patient towards multimodal-multisensor interaction. An interesting point is how the evaluation is done: generalized standard scoring (g-scores) and/or individual scoring (i-scores) and/or time scoring (t-scores) or a combination of those. G-scores are calculated by adding points for successfully solved questions and tasks. Final g-scores take age, sex, and similar factors into account. The digitalization of g-scores is straightforward. I-scores are individual interpretations of the doctor, as in the CDT for example. There is a huge potential to interpret those i-scores in a standardized way in the future. T-scores track the time needed to solve specific sub-tasks. They can be calculated automatically with a digital pen.

Now we focus on the multimodal-multisensor platform (figure 1.8): To capture real-time handwriting input, we employ the Neo Smartpens N2 and M1 [Neosmartpen 2018]. We use the NeoSmartpen SDK that is available on Github for connecting the N2 digital pen and streaming the data using bluetooth. The input is captured as a series of ink samples, which are grouped together forming strokes in pen-up and pen-down events, including timestamps and pen pressure data. The obtained digital ink gets automatically analyzed by the handwriting recognition component, which is based on the Myscript [2018] recognition engine. To distinguish between handwritten input and correction gestures we employ mode detection described by Sonntag et al. [2014a]. Niemann et al. [2018a] use several sensors to capture input and biosignals of the subject during the assessment and are planning to include more in the

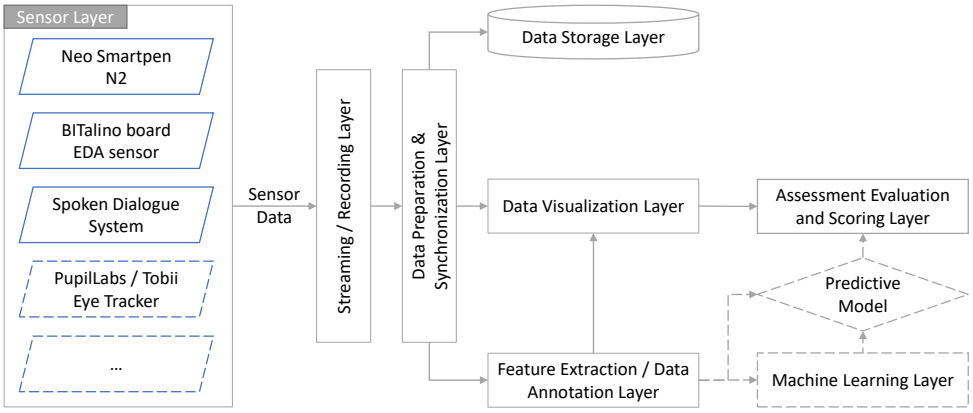

Figure 1.8 Architecture of the multimodal-multisensor platform used for behavior characterization. Dotted lines indicate work in progress.

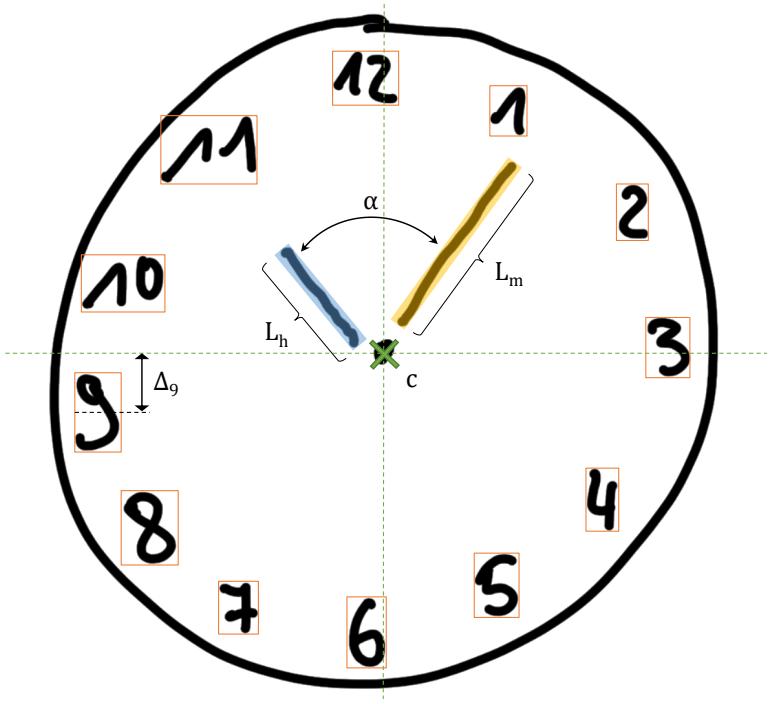

Figure 1.9 Visualization of Multimodal Semantic Features of the Clock Drawing Test.

near future. The *Streaming and Recording Layer* serves as an abstraction layer for the specific hardware components used in the *Sensor Layer*, its main task is to stream and record the raw sensor data. Depending on the exact sensor type different libraries are used to connect the sensor to the overall framework. Synchronization, resampling, and input fusion take place during the *Data Preparation & Synchronization Layer*. Resulting data are visualized (*Data Visualization Layer*) and stored together with the raw input (*Data Storage Layer*) for later usage. The *Feature Extraction / Data Annotation Layer* prepares the handwriting input and EDA sensor data for machine learning tasks (interpretation and late fusion).

Now the focus is on the study design: Werner et al. [2006] used computerized handwriting evaluation to discriminate Mild Alzheimer Disease (AD) and Mild Cognitive Impairment (MCI). They observed that participants with MCI and mild AD spent a significantly longer time with the pen in the air than healthy participants and that all kinematic measures (except for velocity) differ between healthy and impaired participants. Findings by Schroter et al. [2003] suggest that it is possible to distinguish between different forms of cognitive impairment and healthy subjects by analyzing the kinematic aspects of handwriting movements. For the digital pen recording, we prepare the paper on the table, positioned in a comfortable position to the participant, and ask the patients to hold a digital pen in the same hand they hold a normal pen while writing. Inside the tip of the Neo pen an infrared camera recognizes the special microdot pattern printed onto the paper, which is merely visible and therefore similar to normal, white paper. If needed the subject is allowed to fixate the paper using the non-dominant hand and instructed to avoid movement with that hand if possible.

Movement of the hand results in either more or less pressure on the electrodes and, therefore, in noisy data. In order to minimize the chance of false signal peaks from unwanted movement, we suggest putting additional tape on the electrodes to fixate them in-place. The usage of additional cycling gloves has proven to be useful when recording tasks that might include frequent movement of the non-dominant hand (e.g. turning pages). From our experience the amount of movement is highly dependent on the individual subjects. As EDA is a relatively slow signal (latency of about 0.5 to 5 seconds) [Boucsein 2012] in combination with the recovery time of the amplitude, there should be at least 6 seconds between each task of interest. It is advised to run a baseline measurement period between two and four minutes, when the participant is not engaged in any given task [Braithwaite et al. 2013].

Preliminary evaluations of the study design on the CDT, DemTect, and MMSE suggest that, through the automation of these cognitive assessments for dementia, the caregiver's time spent on conducting the tests can be reduced and his or her attention can be shifted from test features while writing (e.g., easy-to-assess completion of input fields) to other more subtle observations. EDA is displayed to the caregiver and helps in interpreting those subtle observations (yet to be quantified).

1.5 Future Directions

Future directions include applications of multimodal-multisensor combinations and virtual reality applications which we will discuss in the rest of this chapter.

1.5.1 Multimodal-Multisensor Combinations

These interfaces combine multiple user input modalities with multiple sensor information (e.g., location, acceleration, proximity, tilt). Sensor-based cues may be used to interpret a user's physical state, health status, mental status, and many other types of information. Sensors may capture biosignals in addition, such as EDA (see case study 3). On the other hand, users may engage in intentional actions when deploying sensor controls, such as hand gestures which are captured by a video sensor.

Simsensei [2014] introduced by [DeVault et al. 2014] is an example of a multimodal-multisensor combination, i.e., a virtual agent-based interface with an additional collection of body sensors. This application should recognize and identify psychological distress from multiple signals in a multimodal dialogue. The mental status subsystem automatically tracks and analyzes in real-time facial expressions, body posture, acoustic features, linguistic patterns and higher-level behavior descriptors (e.g., attention and fidgeting). It is very interesting to mention the two roles the multisensor system has. First, it contributes to the indicator analysis to identify psychological distress from those multiple signals. Just as in case study 2 with digital pen features, these distress indicators can allow the clinician or healthcare provider to make a more informed diagnosis. Second, the sensors' outputs are broadcasted to the other components of the multimodal interface: Sensor outputs assist the virtual human with turn taking, listening feedback, and building rapport by providing appropriate non-verbal feedback.

Other applications are intelligent tutoring systems for educational healthcare. These systems use multimodal presentation of information to allow users (e.g., medical students) with different preferences and abilities to use information in their preferred way. In addition, multisensor processing might include speaker traits. For example, Chatterjee et al. [2015] analyze the most discriminative elements of a speaker's non-verbal behavior that contribute to the perceived credibility or passionateness.

Another example is Kognit [Sonntag 2015], a research project about multisensor input processing to counteract cognitive impairments, based on episodic memory construction through activity recognition. Kognit includes eye tracking sensors for activity recognition and multimodal speech-based dialogue in augmented reality applications. Eye tracking and activity recognition are explored in multiple upcoming research projects, but in Kognit, a robot senses a patient at home and interacts with him or her in a multimodal way. The patient can use speech or a digital pen to communicate with the robot. Multimodal output involves the robot's output from two or more modalities: A head-mounted visual display for the patient is combined with auditory feedback, which is provided as multimodal feedback to the user. Other

important directions of multimodal-multisensor combinations include clinical multimodal-multisensor systems for doctors, where incremental knowledge acquisition, multimodal dialogue constraints, and virtual reality applications are brought together.

1.5.2 Virtual Reality

The design, development and evaluation of virtual reality (VR) systems targets the areas of clinical diagnosis and decision support, clinical assessment, and rehabilitation. VR headsets provide a powerful new tool for future exploration of sensors, for example an Oculus Rift-integrated binocular eye tracking system [Oculus 2018]. In addition, digital pen-based interfaces can be combined with multiple approaches to determining cognitive status through the detection and analysis of subtle behaviors [Davis et al. 2014].

Examples of futuristic VR applications include clinical training [Rizzo and Talbot 2016]. Multimodal dialogues include avatars that responds to pre-selected choices, for example in the context of VR exposure therapy for combat-related posttraumatic stress disorder (PTSD) [Cukor et al. 2016]. Greenleaf [2016] states that significant impact of VR technology will be in the area of clinical medicine and healthcare, mostly because VR can address and ameliorate some of the most difficult problems in healthcare, i.e., ranging from mood disorders such as anxiety and depression to post traumatic stress disorder, addictions, autism, cognitive aging, and physical rehabilitation. VR examples include interactive visualization of shared electronic patient records, previously acquired with a remote tablet device, in a virtual environment. Hand tracking, eye tracking, and vision-based peripheral view monitoring can be integrated. Luxenburger et al. [2016] provide a combination of hand gesture and eye tracking recognition in order to assess whether all regions of a medical image have been explored by the doctor in the VR environment (figure 1.10).

One of the main research questions for the multimodal interaction community is how to match affordances in VR with the medical task domain to (1) stimulate learning and understanding, (2) stimulate cognition, (3) improve overall performance in medical decision support applications. Based on the case study presented in section 1.4.1, Prange et al. [2018a] developed a multimodal VR prototype for the doctor. The system uses a headset google in a remote collaboration setting. In the application scenario, the radiologist starts with a patient examination form by using a tablet with built-in stylus for notes and drawing. The handwritten multi-stroke sketches are transcribed by using handwriting and gesture recognition[1], then analysed and stored based on common medical ontologies [Sonntag et al. 2009]. The doctor then examines the patient records in VR, and he can interact with the 3D MRI medical images of the patient. Ard et al. [2017] present a similar scenario where neurology images are displayed in VR. The end-to-end system of Prange et al. [2018a] provides a GPU-accelerated machine learning model for automated decision support that

[1] http://medicalcps.dfki.de/www/wp-content/uploads/BIRADS-30-seconds.mp4

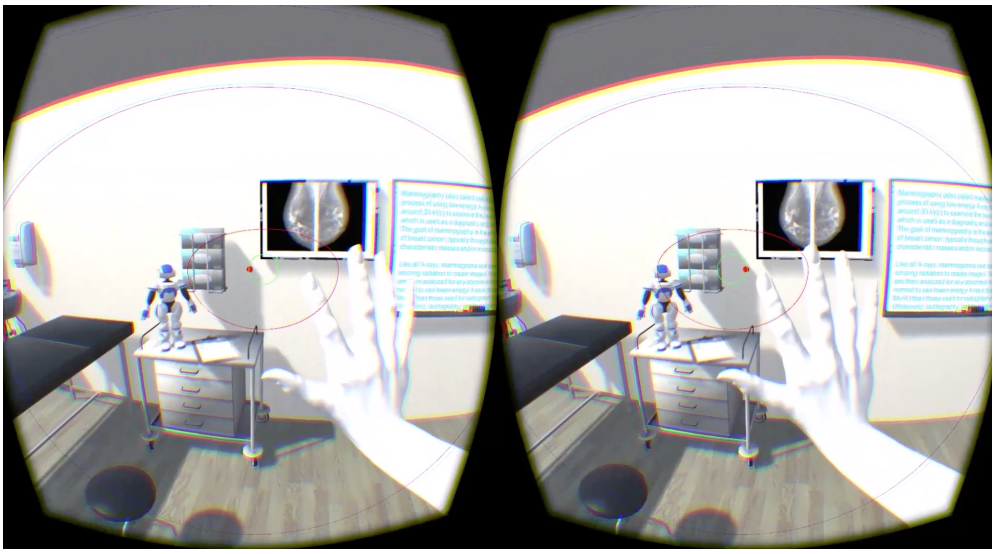

Figure 1.10 Medical remote collaboration using eye gaze and hand gesture input in VR.

computes therapy predictions in real-time. This video[2] shows the complete workflow. The dialogue system supports task-based interaction with the patient data shown on the virtual display (e.g., *"Open the patient file for Gerda Meier."*, *"Show the next page."*), question answering functionality about factoid contents of a patient record (e.g., *"When was the last examination?"*), and the therapy prediction component (*"Which therapy is recommended?"*). This prototype suggests that in future work, it is worth investigating how a VR application, together with dialogue-based therapy prediction, impacts the medical findings process in daily hospital routine, in particular when 3D images can be observed by haptic objects in a natural mapping (figure 1.11). Recent advancement of VR technology for clinical purposes, i.e., added value over traditional diagnosis, decision support, or assessment approaches, may lead to improved immersion effect. Multimodality aspects that can potentially lead to immersion are for example the following: pen, speech, (head) movement, VR controllers for input and a 3D VR scene, 3D image material, animated 3D graphics, and speech for input and output. Since people's object and concept perceptions are multisensor, people are influenced by an array of object affordances (e.g., auditory, tactile) and their visual properties. In addition, the acoustic qualities of a computer voice can influence a user's immersion and engagement. Future work includes additional input modalities such as eye-tracking to improve the multimodal interaction in VR by other physiological sensors.

[2] http://medicalcps.dfki.de/www/wp-content/uploads/KDI_V2_Pro_v04_2.mp4

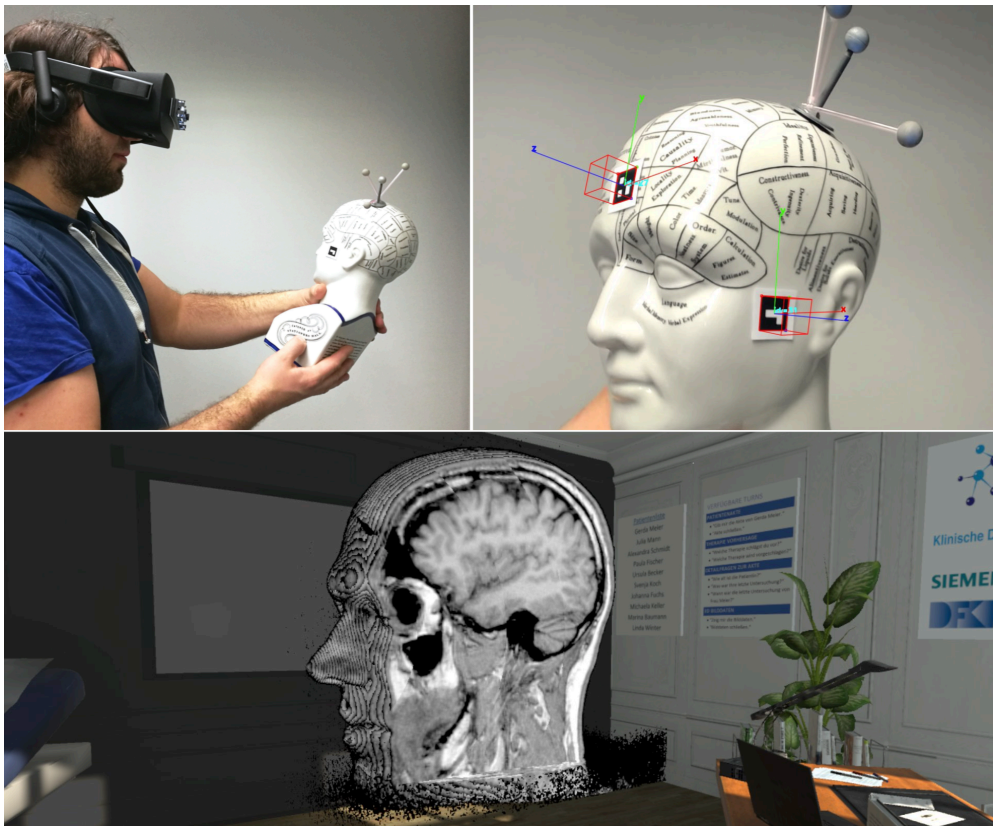

Figure 1.11 Immersion by haptic objects in VR.

1.6 Conclusion

We discussed the trends of multimodal-multisensor interfaces of medical and health systems and emphasized the theoretical foundations of multimodal interfaces and systems in the healthcare domain. We started with a discussion of the background of medical and health systems, defined MCPS, and focused on the distinction of clinical and non-clincial systems, followed by three clinical use case studies. The first study described a multimodal dialogue system in the radiology domain, the second focused on a multisensor digital pen interface for cognitive assessment, and the third described a multimodal-multisensor framework including EDA. Future directions include multimodal-multisensor combinations.

For MCPS, in addition to the specific recommendations of multimodal-multisensor interactions, prototypes will have to go through product lifecycles, including the design, development, distribution, verification, validation, deployment and maintenance of these devices. The exact challenges for real-world MCPS developers, in particular for verification and validation

of such medical interfaces, remain mostly unknown. For example, EDA-based sensor input is easy to capture, but one gets motion artifacts if the hand with the electrodes moves as well as when the state of the hand changes, for instance, from open to close. The changed state of the hand results in either more or less pressure on the electrodes and, therefore, in more or less skin conductance. Future research in medical and health systems should include research on wearable body area networks for continuous monitoring of patients at home, based on wireless sensor networks for healthcare [Alemdar and Ersoy 2010]. Darwish and Hassanien [2011] for example explain the important role of body sensor networks in medicine to minimize the need for caregivers and help the chronically ill and elderly people live an independent life.

In designing future architectures for multimodal-multisensor interfaces for medical and health systems, important insights clearly can be gained from cognitive principles of sensory integration of passive and active input modes. One challenge will be to create human-in-the-loop medical cyber-physical systems that incorporate a broad range of information by data-driven approaches of large multimodal databases. Those data-intensive systems can fuse multiple modalities. These infrastructures are capable of integrating multisensor input, potentially increasing the reliability of a percept through multisensor integration. A further consideration is improved robustness, and hence trust, in future assistive MCPS, where the higher level automatic intent recognition leads to collaborative action with humans in medical care. This also suggests that future research should explore whether individual patients' biomarkers may provide a useful signature for adaptation purposes (of advanced fusion-based multimodal interfaces for example). Such an approach is currently limited by diagnostic tools that are insensitive to changes in behavior future systems should adapt to automatically. Special, new biomarkers are of particular interest, i.e., markers of emotion regulation, social response and social attention. For example, learning representations of affect from speech [Ghosh et al. 2015] can be used in autism detection and treatment and prove that multimodal-multisensor interfaces have medical applications not only in sensomotoric and cognitive intelligence aspects, but also emotional and social intelligence aspects of medicine.

It is beneficial to take a broad perspective. We described the AI perspective that prime applications include clinical decision support, patient monitoring, and automated devices to assist in surgery or patient care. Concerning clinical decision support, advances of multimodal-multisensor interfaces can promise to change the cognitive tasks assigned to human clinicians by cognitive assistants and structured patterns of inference.

1.7 Supplementary Digital Materials: Focus Questions

- Why is WCET (worst-case execution time) an important consideration for medical CPS?
- To support individuals on their personal health we must take a life-time perspective. Why?

- What is the paradigm shift in healthcare provision and how can we manage patients more effectively with multisensor information and multimodal communication technologies?
- Why is it important to focus in the first instance on people with certain risk factors such as cardiovascular risks like high blood pressure or high blood glucose?
- How can emergency response services at home be improved by a multimodal-multisensor interface?
- Why are automatic prevention services so cost-effective? How can multimodal-multisensor interfaces support primary and secondary prevention?
- What information can be extracted from active input modalities to support medical applications based on biosignals?
- How can multimodal-multisensor systems be integrated in daily life situations?
- What adaptation approaches will result in effective multimodal and multisensor interfaces for real-word deployment with patients?

Bibliography

H. Alemdar and C. Ersoy. Oct. 2010. Wireless sensor networks for healthcare: A survey. *Comput. Netw.*, 54(15): 2688–2710. ISSN 1389-1286. http://dx.doi.org/10.1016/j.comnet.2010.05.003. DOI: 10.1016/j.comnet.2010.05.003.

E. Alpaydin. 2018. Classifying multimodal data. In *The Handbook of Multimodal-Multisensor Interfaces*, volume Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 2. Morgan & Claypool Publishers, San Rafael, CA.

T. Ard, D. M. Krum, T. Phan, D. Duncan, R. Essex, M. Bolas, and A. Toga. Mar. 2017. NIVR: Neuro Imaging in Virtual Reality. In *Proceedings of Virtual Reality (VR), 2017 IEEE*, pp. 465–466. IEEE, Los Angeles, CA. ISBN 978-1-5090-6647-6.

T. Baltrusaitis, C. Ahuja, and L.-P. Morency. 2018a. Multimodal machine learning. In *The Handbook of Multimodal-Multisensor Interfaces*, volume Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 1. Morgan & Claypool Publishers, San Rafael, CA.

T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. Morency. May 2018b. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pp. 59–66. DOI: 10.1109/FG.2018.00019.

M. Barz, F. Daiber, D. Sonntag, and A. Bulling. 2018. Error-Aware Gaze-Based Interfaces for Robust Mobile Gaze Interaction. In *Proc. International Symposium on Eye Tracking Research and Applications (ETRA)*.

D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar. Jul 2014. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)*, 33(7): 1123–1131.

S. Bengio, L. Deng, L.-P. Morency, and B. Schuller. 2018. Multidisciplinary challenge topic: Perspectives on predictive power of multimodal deep learning: Surprises and future directions. In *The Handbook of Multimodal-Multisensor Interfaces*, volume Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 14. Morgan & Claypool Publishers, San Rafael, CA.

T. W. Bickmore, L. M. Pfeifer, and B. W. Jack. 2009. Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pp. 1265–1274. ACM, New York, NY, USA. ISBN 978-1-60558-246-7. http://doi.acm.org/10.1145/1518701.1518891. DOI: 10.1145/1518701.1518891.

W. Boucsein. 1992. *Electrodermal activity*. Plenum Press.

W. Boucsein. 2012. *Electrodermal activity*. Springer Science & Business Media.

J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe. 2013. A guide for analysing electrodermal activity (eda) & skin conductance responses (scrs) for psychological experiments. *Psychophysiology*, 49(1): 1017–1034.

R. Canham, S. Smith, and A. Tyrrell. 2000. *Automated scoring of a neuropsychological test: The Rey Osterrieth Complex Figure*, pp. 406–413. IEEE.

P. Carayon. 2011. *Handbook of Human Factors and Ergonomics in Health Care and Patient Safety, Second Edition*. CRC Press.

D. Castro, S. Hickson, V. Bettadapura, E. Thomaz, G. Abowd, H. Christensen, and I. Essa. 2015. Predicting daily activities from egocentric images using deep learning. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ISWC '15, pp. 75–82. ACM, New York, NY, USA. ISBN 978-1-4503-3578-2. http://doi.acm.org/10.1145/2802083.2808398. DOI: 10.1145/2802083.2808398.

M. Cavazza, F. Charles, A. Lindsay, J. Siddle, and G. Georg. 2015. An interactive narrative format for clinical guidelines. *KI - Künstliche Intelligenz*, 29(2): 185–191. ISSN 1610-1987. http://dx.doi.org/10.1007/s13218-015-0354-3. DOI: 10.1007/s13218-015-0354-3.

M. Chatterjee, S. Park, L.-P. Morency, and S. Scherer. 2015. Combining two perspectives on classifying multimodal data for recognizing speaker traits. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pp. 7–14. ACM, New York, NY, USA. ISBN 978-1-4503-3912-4. http://doi.acm.org/10.1145/2818346.2820747. DOI: 10.1145/2818346.2820747.

P. Chaurasia, S. I. McClean, C. D. Nugent, and B. W. Scotney. 2014. A duration-based online reminder system. *Int. J. Pervasive Computing and Communications*, 10(3): 337–366. http://dx.doi.org/10.1108/IJPCC-07-2014-0042. DOI: 10.1108/IJPCC-07-2014-0042.

C. Chen, J. Favre, G. Kurillo, T. P. Andriacchi, R. Bajcsy, and R. Chellappa. 2014a. Camera networks for healthcare, teleimmersion, and surveillance. *IEEE Computer*, 47(5): 26–36. http://doi.ieeecomputersociety.org/10.1109/MC.2014.112. DOI: 10.1109/MC.2014.112.

C. Chen, L. White, T. Kowalewski, R. Aggarwal, C. Lintott, B. Comstock, K. Kuksenok, C. Aragon, D. Holst, and T. Lendvay. 2014b. Crowd-sourced assessment of technical skills: a novel method to evaluate surgical performance. *Journal of Surgical Research*, 187(1): 65 – 71. ISSN 0022-4804. //www.sciencedirect.com/science/article/pii/S0022480413008998. DOI: http://dx.doi.org/10.1016/j.jss.2013.09.024.

Y. Chen, Y. Chen, M. Randriambelonoro, A. Geissbühler, and P. Pu. 2016. Peer influence on the engagement of fitness tracker usage: A diabetic and obesity study. In *2016 IEEE International Conference on Healthcare Informatics, ICHI 2016, Chicago, IL, USA, October 4-7, 2016*, pp. 192–197. http://dx.doi.org/10.1109/ICHI.2016.28. DOI: 10.1109/ICHI.2016.28.

J. F. Cohn, N. Cummins, J. Epps, R. Goecke, J. Joshi, and S. Scherer. 2018. Multimodal assessment of depression and related disorders based on behavioural signals. In *The Handbook of Multimodal-Multisensor Interfaces*, volume Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 12. Morgan & Claypool Publishers, San Rafael, CA.

J. Cukor, M. Gerardi, S. Alley, C. Reist, M. Roy, B. O. Rothbaum, J. Difede, and A. Rizzo. Jan. 2016. Virtual Reality Exposure Therapy for Combat-Related PTSD. In *Posttraumatic Stress Disorder and Related Diseases in Combat Veterans*, pp. 69–83. Springer International Publishing, Cham, Switzerland. ISBN 978-3-319-22984-3 978-3-319-22985-0. http://link.springer.com/10.1007/978-3-319-22985-0_7.

H. P. da Silva, A. Fred, and R. Martins. 2014. Biosignals for everyone. *IEEE Pervasive Computing*, 13(4): 64–71. ISSN 1536-1268. DOI: http://doi.ieeecomputersociety.org/10.1109/MPRV.2014.61.

A. Darwish and A. E. Hassanien. 2011. Wearable and implantable wireless sensor network solutions for healthcare monitoring. *Sensors*, 11(6): 5561–5595. ISSN 1424-8220. http://www.mdpi.com/

1424-8220/11/6/5561. DOI: 10.3390/s110605561.

R. Davis, D. Libon, R. Au, D. Pitman, and D. Penney. 2014. Think: Inferring cognitive status from subtle behaviors. In *Innovative Applications of Artificial Intelligence (IAAI)*. http://www.aaai.org/ocs/index.php/IAAI/IAAI14/paper/view/8626.

F. de Man, S. Greuters, C. Boer, D. Veerman, and S. Loer. 2013. Intra-operative monitoring, many alarms with minor impact. *Anaesthesia*, 68.

D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet, G. Lucas, S. Marsella, F. Morbini, A. Nazarian, S. Scherer, G. Stratou, A. Suri, D. Traum, R. Wood, Y. Xu, A. Rizzo, and L.-P. Morency. 2014. Simsensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '14, pp. 1061–1068. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC. ISBN 978-1-4503-2738-1. http://dl.acm.org/citation.cfm?id=2617388.2617415.

S. K. D'Mello, N. Bosch, and H. Chen. 2018. Multimodal-multisensor affect detection. In *The Handbook of Multimodal-Multisensor Interfaces*, volume Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 6. Morgan & Claypool Publishers, San Rafael, CA.

F. Eyben, F. Weninger, F. Gross, and B. Schuller. 2013. Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, pp. 835–838. ACM, New York, NY, USA. ISBN 978-1-4503-2404-5. http://doi.acm.org/10.1145/2502081.2502224. DOI: 10.1145/2502081.2502224.

M. Folstein, S. Folstein, and P. McHugh. Nov 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3): 189–198.

E. S. Ford, M. M. Bergmann, J. Kröger, A. Schienkiewitz, C. Weikert, and H. Boeing. 2009. Healthy living is the best revenge: Findings from the european prospective investigation into cancer and nutrition?potsdam study. *Archives of Internal Medicine*, 169(15): 1355–1362. +http://dx.doi.org/10.1001/archinternmed.2009.237. DOI: 10.1001/archinternmed.2009.237.

M. Freedman, L. Leach, E. Kaplan, G. Winocur, K. Shulman, and D. Delis. 1994. *Clock Drawing: A Neuropsychological Analysis*. Oxford University Press. ISBN 9780198022565.

G. Friedland and M. Tschantz. 2018. Privacy concerns of multimodal sensor systems. In S. Oviatt, B. Schuller, P. R. Cohen, D. Sonntag, G. Potamianos, and A. Krüger, eds., *The Handbook of Multimodal-Multisensor Interfaces, Volume 3*. Association for Computing Machinery and Morgan & Claypool, New York, NY, USA.

M. Garbarino, M. Lai, D. Bender, R. Picard, and S. Tognetti. Nov. 2014. Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *2014 EAI 4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth)*, pp. 39–42. DOI: 10.1109/MOBIHEALTH.2014.7015904.

G. Gatterer, P. Fischer, M. Simanyi, and W. Danielczyk. 1989. The A-K-T ("Alters-Konzentrations-Test") a new psychometric test for geriatric patients. *Funct. Neurol.*, 4(3): 273–276.

J. Gelissen and D. Sonntag. 2015. Special issue on health and wellbeing. *KI - Künstliche Intelligenz*, 29(2): 111–113. ISSN 1610-1987. http://dx.doi.org/10.1007/s13218-015-0360-5. DOI: 10.1007/s13218-015-0360-5.

GenieMD, 2017. GenieMD. http://geniemd.com. 2018-03-21.

S. Ghosh, E. Laksana, L. Morency, and S. Scherer. 2015. Learning representations of affect from speech. *CoRR*, abs/1511.04747. http://arxiv.org/abs/1511.04747.

D. Graus, P. N. Bennett, R. W. White, and E. Horvitz. 2016. Analyzing and predicting task reminders. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, UMAP '16, pp. 7–15. ACM, New York, NY, USA.

W. Greenleaf. 2016. How VR Technology Will Transform Healthcare. In *ACM SIGGRAPH 2016 VR Village*, SIGGRAPH '16, pp. 5:1–5:2. ACM, New York, NY, USA. ISBN 978-1-4503-4377-0. http://doi.acm.org/10.1145/2929490.2956569. DOI: 10.1145/2929490.2956569.

S. Haddadin, A. D. Luca, and A. Albu-Schäffer. Dec 2017. Robot collisions: A survey on detection, isolation, and identification. *IEEE Transactions on Robotics*, 33(6): 1292–1312. ISSN 1552-3098. DOI: 10.1109/TRO.2017.2723903.

P. J Wagner, H. Wortzel, K. Frey, C. Alan Anderson, and D. Arciniegas. 2011. Clock-drawing performance predicts inpatient rehabilitation outcomes after traumatic brain injury. *The Journal of neuropsychiatry and clinical neurosciences*, 23: 449–53.

M. Jacob, Y.-T. Li, G. Akingba, and J. P. Wachs. 2012. Gestonurse: a robotic surgical nurse for handling surgical instruments in the operating room. *Journal of Robotic Surgery*, 6(1): 53–63. ISSN 1863-2491. http://dx.doi.org/10.1007/s11701-011-0325-0. DOI: 10.1007/s11701-011-0325-0.

M. G. Jacob and J. P. Wachs. Jan. 2014. Context-based hand gesture recognition for the operating room. *Pattern Recogn. Lett.*, 36: 196–203. ISSN 0167-8655. http://dx.doi.org/10.1016/j.patrec.2013.05.024. DOI: 10.1016/j.patrec.2013.05.024.

E. Kalbe, J. Kessler, P. Calabrese, R. Smith, A. P. Passmore, M. Brand, and R. Bullock. Feb 2004. DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *International Journal of Geriatric Psychiatry*, 19(2): 136–143.

G. Keren, A. E.-D. Mousa, O. Pietquin, S. Zafeiriou, and B. Schuller. 2018. Deep learning for multisensorial and multimodal interaction. In *The Handbook of Multimodal-Multisensor Interfaces*, volume Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 4. Morgan & Claypool Publishers, San Rafael, CA.

H. Kurniawan, A. V. Maslov, and M. Pechenizkiy. June 2013. Stress detection from speech and galvanic skin response signals. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pp. 209–214. DOI: 10.1109/CBMS.2013.6627790.

A. Lally, S. Bagchi, M. Barborak, D. W. Buchanan, J. Chu-Carroll, D. A. Ferrucci, M. R. Glass, A. Kalyanpur, E. T. Mueller, J. W. Murdock, S. Patwardhan, and J. M. Prager. 2017. Watsonpaths: Scenario-based question answering and inference over unstructured information. *AI Magazine*, 38(2): 59–76. https://www.aaai.org/ojs/index.php/aimagazine/article/view/2715.

C. P. Langlotz. 2006. Radlex: A new method for indexing online educational materials. *RadioGraphics*, 26: 1595–1597. http://radiographics.rsnajnls.org/cgi/content/full/26/6/1595. DOI: doi:10.1148/rg.266065168.

I. Lee, O. Sokolsky, S. Chen, J. Hatcliff, E. Jee, B. Kim, A. King, M. Mullen-Fortino, S. Park, A. Roederer, and K. Venkatasubramanian. Jan 2012. Challenges and research directions in medical cyber-physical systems. *Proceedings of the IEEE*, 100(1): 75–90.

C. Lisetti, R. Amini, and U. Yasavur. 2015. Now all together: Overview of virtual health assistants emulating face-to-face health interview experience. *KI - Künstliche Intelligenz*, 29(2): 161–172. ISSN

1610-1987. http://dx.doi.org/10.1007/s13218-015-0357-0. DOI: 10.1007/s13218-015-0357-0.

S. Lupien, F. Maheu, M. Tu, A. Fiocco, and T. Schramek. 2007. The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and Cognition*, 65(3): 209 – 237. ISSN 0278-2626. http://www.sciencedirect.com/science/article/pii/S0278262607000322. DOI: https://doi.org/10.1016/j.bandc.2007.02.007.

A. Luxenburger, A. Prange, M. M. Moniri, and D. Sonntag. 2016. Medicalvr: Towards medical remote collaboration using virtual reality. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, pp. 321–324. ACM, New York, NY, USA. ISBN 978-1-4503-4462-3. http://doi.acm.org/10.1145/2968219.2971392. DOI: 10.1145/2968219.2971392.

S. Luz and B. Kane. 2009. Classification of patient case discussions through analysis of vocalisation graphs. In *Proceedings of the 2009 International Conference on Multimodal Interfaces*, ICMI-MLMI '09, pp. 107–114. ACM, New York, NY, USA. ISBN 978-1-60558-772-1. http://doi.acm.org/10.1145/1647314.1647334. DOI: 10.1145/1647314.1647334.

J.-C. Martin, C. Clavel, M. Courgeon, M. Ammi, M.-A. Amorim, Y. Tsalamlal, and Y. Gaffary. 2018. How do users perceive multimodal expressions of affects? In *The Handbook of Multimodal-Multisensor Interfaces*, volume Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 9. Morgan & Claypool Publishers, San Rafael, CA.

U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher. 2006. Activity recognition and monitoring using multiple sensors on different body positions. In *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*.

K. R. McKeown, S.-F. Chang, J. Cimino, S. Feiner, C. Friedman, L. Gravano, V. Hatzivassiloglou, S. Johnson, D. A. Jordan, J. L. Klavans, A. Kushniruk, V. Patel, and S. Teufel. 2001. Persival, a system for personalized search and summarization over multimedia healthcare information. In *Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '01, pp. 331–340. ACM, New York, NY, USA. ISBN 1-58113-345-6. http://doi.acm.org/10.1145/379437.379722. DOI: 10.1145/379437.379722.

MedicalCPS, 2018. Medical cyber-physical systems. http://dfki.de/MedicalCPS/. 2018-08-22.

G. Mehlmann, K. Janowski, and E. André. 2016. Modeling grounding for interactive social companions. *KI - Künstliche Intelligenz*, 30(1): 45–52. ISSN 1610-1987. http://dx.doi.org/10.1007/s13218-015-0397-5. DOI: 10.1007/s13218-015-0397-5.

J. L. Mejino, D. L. Rubin, and J. F. Brinkley. 2008. FMA-RadLex: An application ontology of radiological anatomy derived from the foundational model of anatomy reference ontology. In *Proc. of AMIA Symposium*, pp. 465–469. http://stanford.edu/~rubin/pubs/097.pdf.

M. Möller, S. Regel, and M. Sintek. June 2009. Radsem: Semantic annotation and retrieval for medical images. In *Proc. of The 6th Annual European Semantic Web Conference (ESWC2009)*. http://www.manuelm.org/publications/wp-content/uploads/2009/02/eswc2009.pdf.

M. Möller, M. Sintek, R. Biedert, P. Ernst, A. Dengel, and D. Sonntag. 2010. Representing the international classification of diseases version 10 in OWL. In J. Filipe and J. L. G. Dietz, eds., *KEOD 2010 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Valencia, Spain, October 25-28, 2010*, pp. 50–59. SciTePress. ISBN 978-989-8425-29-4.

N. Moraveji and C. Soesanto. 2012. Towards stress-less user interfaces: 10 design heuristics based on the psychophysiology of stress. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pp. 1643–1648. ACM, New York, NY, USA. ISBN 978-1-4503-1016-1. http://doi.acm.org/10.1145/2212776.2223686. DOI: 10.1145/2212776.2223686.

J. Morris, R. Mohs, H. Rogers, G. Fillenbaum, and A. Heyman. 1988. Consortium to establish a registry for Alzheimer's disease (CERAD) clinical and neuropsychological assessment of Alzheimer's disease. *Psychopharmacol Bull.*, 24(4): 641–52.

A. Mourão and F. Martins. 2013. Novamedsearch: A multimodal search engine for medical case-based retrieval. In *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval*, OAIR '13, pp. 223–224. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, Paris, France, France. ISBN 978-2-905450-09-8. http://dl.acm.org/citation.cfm?id=2491748.2491798.

S. Munir, J. A. Stankovic, C.-J. M. Liang, and S. Lin. 2013. Cyber physical system challenges for human-in-the-loop control. In *Presented as part of the 8th International Workshop on Feedback Computing*. USENIX, Berkeley, CA. https://www.usenix.org/conference/feedbackcomputing13/workshop-program/presentation/Munir.

Myscript, 2018. Myscript SDK product. https://www.myscript.com. 2018-11-05.

Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow. 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4): 695–699.

Neosmartpen, 2018. Neo N2 smartpen product. https://www.neosmartpen.com. 2018-11-05.

J. Nielsen and R. L. Mack, eds. 1994. *Usability Inspection Methods*. John Wiley & Sons, Inc., New York, NY, USA. ISBN 0-471-01877-5.

M. Niemann, A. Prange, and D. Sonntag. 2018a. Towards a multimodal multisensory cognitive assessment framework. In J. Hollmén, C. McGregor, P. Soda, and B. Kane, eds., *31st IEEE International Symposium on Computer-Based Medical Systems, CBMS 2018, Karlstad, Sweden, June 18-21, 2018*, pp. 24–29. IEEE Computer Society. ISBN 978-1-5386-6060-7. https://doi.org/10.1109/CBMS.2018.00012. DOI: 10.1109/CBMS.2018.00012.

M. Niemann, A. Prange, and D. Sonntag. 2018b. Towards a multimodal multisensory cognitive assessment framework. In *IEEE 31th International Symposium on Computer-Based Medical Systems (CBMS)*.

Oculus, 2018. Oculus Rift product. https://www.oculus.com. 2018-11-05.

S. Oviatt. 1999. Mutual disambiguation of recognition errors in a multimodel architecture. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pp. 576–583. ACM, New York, NY, USA. ISBN 0-201-48559-1. http://doi.acm.org/10.1145/302979.303163. DOI: 10.1145/302979.303163.

S. Oviatt and P. Cohen. Mar. 2000. Perceptual user interfaces: Multimodal interfaces that process what comes naturally. *Commun. ACM*, 43(3): 45–53. ISSN 0001-0782. http://doi.acm.org/10.1145/330534.330538. DOI: 10.1145/330534.330538.

S. Oviatt and P. R. Cohen. 2015. *The Paradigm Shift to Multimodality in Contemporary Computer Interfaces*. Morgan & Claypool. ISBN 9781627057523. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7087855. DOI: 10.2200/S00636ED1V01Y201503HCI030.

S. Oviatt, J. F. Grafsgaard, L. Chen, and X. Ochoa. 2018a. Multimodal learning analytics: Assessing learners' mental state during the process of learning. In *The Handbook of Multimodal-Multisensor Interfaces*, volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 11. Morgan & Claypool Publishers, San Rafael, CA.

S. Oviatt, B. Schuller, P. R. Cohen, D. Sonntag, G. Potamianos, and A. Krüger, eds. 2018b. *The Handbook of Multimodal-Multisensor Interfaces*, volume 2. Association for Computing Machinery and Morgan & Claypool, New York, NY, USA.

S. L. Oviatt, K. Hang, J. Zhou, K. Yu, and F. Chen. 2018c. Dynamic handwriting signal features predict domain expertise. *TiiS*, 8(3): 18:1–18:21. http://doi.acm.org/10.1145/3213309. DOI: 10.1145/3213309.

Y. Panagakis, O. Rudovic, and M. Pantic. 2018. Learning for multi-modal and context-sensitive interfaces. In *The Handbook of Multimodal-Multisensor Interfaces*, volume Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 3. Morgan & Claypool Publishers, San Rafael, CA.

A. Pecchinenda. 1996. The affective significance of skin conductance activity during a difficult problem-solving task. *Cognition & Emotion*, 10(5): 481–504.

R. W. Picard, M. Migliorini, C. Caborni, F. Onorati, G. Regalia, D. Friedman, and O. Devinsky. 2017. Wrist sensor reveals sympathetic hyperactivity and hypoventilation before probable sudep. *Neurology*. ISSN 0028-3878. http://n.neurology.org/content/early/2017/07/12/WNL.0000000000004208. DOI: 10.1212/WNL.0000000000004208.

A. Prange, I. P. Sandrala, M. Weber, and D. Sonntag. 2015. Robot companions and smartpens for improved social communication of dementia patients. In *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion*, IUI Companion '15, pp. 65–68. ACM, New York, NY, USA. ISBN 978-1-4503-3308-5. http://doi.acm.org/10.1145/2732158.2732174. DOI: 10.1145/2732158.2732174.

A. Prange, M. Barz, and D. Sonntag. 2018a. Medical 3d images in multimodal virtual reality. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, IUI'18, pp. 19:1–19:2. ACM, New York, NY, USA. ISBN 978-1-4503-5571-1. http://doi.acm.org/10.1145/3180308.3180327. DOI: 10.1145/3180308.3180327.

A. Prange, M. Barz, and D. Sonntag, 2018b. A categorisation and implementation of digital pen features for behaviour characterisation.

RadSpeech, 2011. RadSpeech Project. https://www.dfki.de/RadSpeech. 2018-10-31.

R. Reitan. 1992. *Trail Making Test*. Reitan Neuropsychology Laboratory.

S. Rick, A. Calvitti, Z. Agha, and N. Weibel. 2015. Eyes on the clinic: Accelerating meaningful interface analysis through unobtrusive eye tracking. In *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '15, pp. 213–216. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium. ISBN 978-1-63190-045-7. http://dl.acm.org/citation.cfm?id=2826165.2826197.

A. Rizzo and T. Talbot. Jan. 2016. Virtual Reality Standardized Patients for Clinical Training. In *The Digital Patient*, pp. 255–272. John Wiley & Sons, Inc, Hoboken, NJ. ISBN 978-1-118-95278-8 978-1-118-95275-7. http://doi.wiley.com/10.1002/9781118952788.ch18.

C. Saitis and K. Kalimeri. 2016. Identifying urban mobility challenges for the visually impaired with mobile monitoring of multimodal biosignals. In *International Conference on Universal Access in Human-Computer Interaction*, pp. 616–627. Springer.

M. Samwald, A. Jentzsch, C. Bouton, C. Kallesøe, E. L. Willighagen, J. Hajagos, M. S. Marshall, E. Prud'hommeaux, O. Hassanzadeh, E. Pichler, and S. Stephens. 2011. Linked open drug data for pharmaceutical research and development. *J. Cheminformatics*, 3: 19. http://dx.doi.org/10.1186/1758-2946-3-19. DOI: 10.1186/1758-2946-3-19.

Y. Sawamoto, Y. Koyama, Y. Hirano, S. Kajita, K. Mase, K. Katsuyama, and K. Yamauchi. 2007. Extraction of important interactions in medical interviews using nonverbal information. In *Proceedings of the 9th International Conference on Multimodal Interfaces*, ICMI '07, pp. 82–85. ACM, New York, NY, USA. ISBN 978-1-59593-817-6. http://doi.acm.org/10.1145/1322192.1322209. DOI: 10.1145/1322192.1322209.

S. Scherer, G. Stratou, and L.-P. Morency. 2013. Audiovisual behavior descriptors for depression assessment. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ICMI '13, pp. 135–140. ACM, New York, NY, USA. ISBN 978-1-4503-2129-7. http://doi.acm.org/10.1145/2522848.2522886. DOI: 10.1145/2522848.2522886.

A. Schroter, R. Mergl, K. Burger, H. Hampel, H. J. Moller, and U. Hegerl. 2003. Kinematic analysis of handwriting movements in patients with Alzheimer's disease, mild cognitive impairment, depression and healthy subjects. *Dementia and Geriatric Cognitive Disorders*, 15(3): 132–142.

B. Schuller. 2018. Multimodal user state & trait recognition: An overview. In *The Handbook of Multimodal-Multisensor Interfaces*, volume Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 5. Morgan & Claypool Publishers, San Rafael, CA.

Sesame, 2017. RDF4J sesame. http://rdf4j.org. 2017-05-22.

A. Shademan, R. S. Decker, J. D. Opfermann, S. Leonard, A. Krieger, and P. C. W. Kim. 2016. Supervised autonomous robotic soft tissue surgery. *Science Translational Medicine*, 8(337): 337ra64–337ra64. ISSN 1946-6234. http://stm.sciencemag.org/content/8/337/337ra64. DOI: 10.1126/scitranslmed.aad9398.

H. P. D. Silva, S. Fairclough, A. Holzinger, R. Jacob, and D. Tan. Jan. 2015. Introduction to the special issue on physiological computing for human-computer interaction. *ACM Trans. Comput.-Hum. Interact.*, 21(6): 29:1–29:4. ISSN 1073-0516. http://doi.acm.org/10.1145/2688203. DOI: 10.1145/2688203.

Simsensei, 2014. Project description. http://ict.usc.edu/prototypes/simsensei/. 2017-01-16.

D. Sonntag. 2015. Kognit: Intelligent cognitive enhancement technology by cognitive models and mixed reality for dementia patients. In *Proceedings of the AAAI Fall Symposium Series*. http://www.aaai.org/ocs/index.php/FSS/FSS15/paper/view/11702.

D. Sonntag. 2016. Medical cyber-physical systems. In *Cyber-Physical system design with sensor networking technologies*, Control, Robotics and Sensors, pp. 311–333. Institution of Engineering and Technology.

D. Sonntag. 2017. Interakt - A multimodal multisensory interactive cognitive assessment tool. *CoRR*, abs/1709.01796. http://arxiv.org/abs/1709.01796.

D. Sonntag and M. Möller. 2010. A multimodal dialogue mashup for medical image semantics. In *Proceedings of the 15th International Conference on Intelligent User Interfaces*, IUI '10, pp. 381–

384. ACM, New York, NY, USA. ISBN 978-1-60558-515-4. http://doi.acm.org/10.1145/1719970.1720036. DOI: 10.1145/1719970.1720036.

D. Sonntag and D. Porta. 2014. Intelligent semantic mediation, knowledge acquisition and user interaction. In Wahlster et al. [2014], pp. 179–189.

D. Sonntag, P. Wennerberg, P. Buitelaar, and S. Zillner. 2009. Pillars of ontology treatment in the medical domain. *J. Cases on Inf. Techn.*, 11(4): 47–73.

D. Sonntag, N. Reithinger, G. Herzog, and T. Becker. 2010a. A discourse and dialogue infrastructure for industrial dissemination. In *Proceedings of the Second International Conference on Spoken Dialogue Systems for Ambient Environments*, IWSDS'10, pp. 132–143. Springer-Verlag, Berlin, Heidelberg. ISBN 3-642-16201-0, 3-642-16201-5. http://dl.acm.org/citation.cfm?id=1925948.1925961.

D. Sonntag, C. Weihrauch, O. Jacobs, and D. Porta. 4 2010b. Theseus ctc-wp4 usability guidelines for use case applications. Technical report, Bundesministerium für Wirtschaft und Technologie.

D. Sonntag, C. Schulz, C. Reuschling, and L. Galarraga. 2012. Radspeech's mobile dialogue system for radiologists. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces*, IUI '12, pp. 317–318. ACM, New York, NY, USA. ISBN 978-1-4503-1048-2. http://doi.acm.org/10.1145/2166966.2167031. DOI: 10.1145/2166966.2167031.

D. Sonntag, M. Weber, A. Cavallaro, and M. Hammon. 2014a. Integrating digital pens in breast imaging for instant knowledge acquisition. *AI Magazine*, 35(1): 26–37.

D. Sonntag, S. Zillner, P. Ernst, C. Schulz, M. Sintek, and P. Dankerl. 2014b. Mobile radiology interaction and decision support systems of the future. In Wahlster et al. [2014], pp. 371–382.

D. Sonntag, V. Tresp, S. Zillner, A. Cavallaro, M. Hammon, A. Reis, A. P. Fasching, M. Sedlmayr, T. Ganslandt, H.-U. Prokosch, K. Budde, D. Schmidt, C. Hinrichs, T. Wittenberg, P. Daumke, and G. P. Oppelt. 2015. The clinical data intelligence project. *Informatik-Spektrum*, pp. 1–11. ISSN 1432-122X. http://dx.doi.org/10.1007/s00287-015-0913-x. DOI: 10.1007/s00287-015-0913-x.

W. Souillard-Mandar, R. Davis, C. Rudin, R. Au, D. J. Libon, R. Swenson, C. C. Price, M. Lamar, and D. L. Penney. mar 2016. Learning classification models of cognitive conditions from subtle behaviors in the digital Clock Drawing Test. *Machine Learning*, 102(3): 393–441. ISSN 1573-0565. https://doi.org/10.1007/s10994-015-5529-5. DOI: 10.1007/s10994-015-5529-5.

J. C. Sriram, M. Shin, T. Choudhury, and D. Kotz. 2009. Activity-aware ecg-based patient authentication for remote health monitoring. In *Proceedings of the 2009 International Conference on Multimodal Interfaces*, ICMI-MLMI '09, pp. 297–304. ACM, New York, NY, USA. ISBN 978-1-60558-772-1. http://doi.acm.org/10.1145/1647314.1647378. DOI: 10.1145/1647314.1647378.

Taurus, 2017. Taurus. https://www.youtube.com/watch?v=cqBm97jBvuY. 2015-05-06.

A. Tobergte, P. Helmer, U. Hagn, P. Rouiller, S. Thielmann, S. Grange, A. Albu-Schäffer, F. Conti, and G. Hirzinger. 2011. The sigma.7 haptic interface for mirosurge: A new bi-manual surgical console. In *IROS*, pp. 3023–3030. IEEE. ISBN 978-1-61284-454-1. http://dblp.uni-trier.de/db/conf/iros/iros2011.html#TobergteHHRTGACH11.

TUG, 2017. Tug robots in healthcare. http://www.aethon.com/tug/tughealthcare/. 2017-05-22.

G. Turchetti, I. Palla, F. Pierotti, and A. Cuschieri. 2012. Economic evaluation of da vinci-assisted robotic surgery: a systematic review. *Surgical Endoscopy*, 26(3): 598–606. ISSN 1432-2218. http://dx.doi.org/10.1007/s00464-011-1936-2. DOI: 10.1007/s00464-011-1936-2.

M. Valstar. 2014. Automatic behaviour understanding in medicine. In *Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research Including Business Opportunities and Challenges*, RFMIR '14, pp. 57–60. ACM, New York, NY, USA. ISBN 978-1-4503-0615-7. http://doi.acm.org/10.1145/2666253.2666260. DOI: 10.1145/2666253.2666260.

Verbsurgical, 2017. Verbsurgical. http://www.verbsurgical.com. 2015-05-22.

J. Wagner and E. André. 2018. Real-time sensing of affect and social signals in a multimodal framework: a practical approach. In *The Handbook of Multimodal-Multisensor Interfaces*, volume Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 8. Morgan & Claypool Publishers, San Rafael, CA.

W. Wahlster, H. Grallert, S. Wess, H. Friedrich, and T. Widenka, eds. 2014. *Towards the Internet of Services: The THESEUS Research Program*. Cognitive Technologies. Springer.

T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman. 2008. Aligning temporal data by sentinel events: Discovering patterns in electronic health records. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pp. 457–466. ACM, New York, NY, USA. ISBN 978-1-60558-011-1. http://doi.acm.org/10.1145/1357054.1357129. DOI: 10.1145/1357054.1357129.

N. Weibel, C. Emmenegger, J. Lyons, R. Dixit, L. L. Hill, and J. D. Hollan. 2013. Interpreter-mediated physician-patient communication: Opportunities for multimodal healthcare interfaces. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '13, pp. 113–120. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium. ISBN 978-1-936968-80-0. http://dx.doi.org/10.4108/icst.pervasivehealth.2013.252026. DOI: 10.4108/icst.pervasivehealth.2013.252026.

G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda, and A. J. Schreiber. 2016. Smartwatch-based activity recognition: A machine learning approach. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 426–429.

P. Werner, S. Rosenblum, G. Bar-On, J. Heinik, and A. Korczyn. 2006. Handwriting process variables discriminating mild alzheimer's disease and mild cognitive impairment. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 61(4): P228–P236.

K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman. 2011. Lifeflow: Visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 1747–1756. ACM, New York, NY, USA. ISBN 978-1-4503-0228-9. http://doi.acm.org/10.1145/1978942.1979196. DOI: 10.1145/1978942.1979196.

A. D. Wood and J. A. Stankovic. Jan. 2008. Human in the loop: Distributed data streams for immersive cyber-physical systems. *SIGBED Rev.*, 5(1): 20:1–20:2. ISSN 1551-3688. http://doi.acm.org/10.1145/1366283.1366303. DOI: 10.1145/1366283.1366303.

J. Zhai and A. Barreto. 2006. Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. ISBN 1-4244-0032-5.

J. Zhou, K. Yu, F. Chen, Y. Wang, and S. Z. Arshad. 2018. Multimodal behavioural and physiological signals as indicators of cognitive load. In *The Handbook of Multimodal-Multisensor Interfaces*,

volume Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition, chapter 10. Morgan & Claypool Publishers, San Rafael, CA.