Situation Aware Interaction with Multi-modal Business Applications in Smart Environments

Mario Aehnelt[1], Sebastian Bader[2], Gernot Ruscher[2], Frank Krüger[2], Bodo Urban[1], and Thomas Kirste[2]

[1] Fraunhofer IGD, Joachim-Jungius-Strasse 11, 18059 Rostock, Germany
{mario.aehnelt,bodo.urban}@igd-r.fraunhofer.de
[2] University of Rostock, Albert-Einstein-Strasse 22, 18059 Rostock, Germany
{sebastian.bader,gernot.ruscher,frank.krueger2, thomas.kirste}@uni-rostock.de

Abstract. A consistent user experience in combination with proactive assistance may improve the user performance while interacting with heterogeneous data sources as e.g., occurring in business decision making. We describe our approach which is based on inferring the user intentions from sensory inputs, providing a situation aware information assistance, and controlling the environment proactively by anticipating future goals. Our system has been realized within a smart meeting room and has in parts been evaluated. In this paper, we describe the core ideas underlying our approach and report on first findings from the evaluation.

Keywords: intelligent environment, proactive assistance, self-explanation, information assistance, interaction design, usability.

1 Introduction

Business decisions require substantial and interconnected information. However, in real world's environments - as we face them in today's manufacturing industries for example - this information is derived from heterogeneous company resources and systems, e.g. enterprise resource planning systems (ERP) or manufacturing execution systems (MES). Thus, presentation of and interaction with required information follows diverse strategies and metaphors used by underlying devices, applications and user interfaces. The growing use of novel mobile or display devices with completely different interaction metaphors like touch gestures or voice recognition raises further problems. It leads to alternating user experiences and influences the efficiency of decision making processes [1].

In this paper we explore how far recent approaches from HCI and intention recognition can be used to improve the efficiency of business processes. It is our goal to increase process reliability and decrease process costs through improving homogeneity and quality of interacting with data and information using business applications embedded in smart environments. In order to work with realistic business

S. Yamamoto (Ed.): HIMI/HCII 2013, Part III, LNCS 8018, pp. 413–422, 2013.

scenarios we start our exploration by illustrating a specific use case which will serve as basis for our further consideration.

2 Use Case: Smart Business Applications in Manufacturing Industries

In manufacturing industries we find a vital demand for decision making on all company levels. Even on the shop floor a growing amount of structured and unstructured data requires aggregation, interpretation and decision making to guarantee a timely and qualitative delivery of industrial goods. This data reaches from work order information and product specifications to customer data and correspondence as well as information on the current state of production, issues and work plans. Within this general scenario we want to address decision making processes on operational level where the monitoring and controlling of all manufacturing and assembling activities takes place. Our use case here involves different roles: assembling team manager, technical supervisor, constructing engineer and purchaser. In a daily routine they watch over the current and planned state of production and assembling in order to identify upcoming problems and solve them through an early adoption of prior plans. Their decision making here will build the theoretical and practical testbed for our further work.

We transfer the described meeting into a smart control room with a multi-touch table surrounded by an instrumented multi-display environment including a position sensor system. There the assembling team manager is meeting with his supervisor in order to discuss the team's recent progress and problems. They use their private laptops and the multi-touch table to work through the plans and work sheets. In discussion they face a critical issue with a pump in a cooler engine. One of the purchasers is immediately needed to check a replacement. As soon as he enters, the small screens are not sufficient anymore and the presentation is moved to a larger public display next to the table. It shows a detailed model of the pump to replace in parallel to the current work plan. To discuss further technical issues, a constructing engineer is required. He is connected via Skype remote conferencing and shown on a public display to interact with all the other team members simultaneously. In order to start an individual discussion on technical details and on pricing the purchaser moves to another room corner. This movement is recognized by the smart control room and as a result the video chat is transferred onto another screen in this corner. Meanwhile, team leader and his supervisor are discussing further work plans on the table display. As soon as purchaser and constructing engineer have finished their discussion, the room transfers their chat back to the public display. The supervisor can finally estimate delays based on his team's feedback and decide next steps in order to solve the issue. After leaving the smart control room is set back. All described steps of the team meeting can be repeated in arbitrary order.

This setup has been implemented in the *SmartLab* as described in [2] using the *Plant@Hand* system which provides monitoring, planning and documentation functions for manufacturing or maintenance companies (see Figure 1).

Fig. 1. Plant@Hand multi-touch user interface with linked information and in use during a team discussion

3 Preliminaries

Some preliminary evaluation and work results with respect to specific aspects of our approach are discussed below.

3.1 Dynamic and Heterogeneous Smart Environments

For our research, we do not assume any fixed setup of the environment. On the contrary, it is assumed to be a dynamically changing ensemble of heterogeneous devices and services. Our middleware [3] allows connecting all components dynamically. In addition it realizes the idea of *goal-based interaction* as described in [4]. Instead of a command on a given device, it suffices to specify a goal. The component itself will realize this goal. A possible goal for a projector is "*show Input VGA 1*", depending on the current state different actions are necessary to reach this goal (switch power, mute, input source, etc.), but these details are handled by the device itself.

3.2 Multi-modal Interaction

Based on the heterogeneous and dynamically changing ensemble of devices where we cannot rely on any central component, we developed a multi-modal interaction system [2]. *Dialogues* can be described as flows of interactions between user and system, composed of actions, choice, loops and sequences, and entirely independent of the equipment of the underlying environment. User-system interaction is realized by *Interaction Components* which encapsulate different modalities such as GUI, speech, and also sensor data. User-initiated interaction is handled by components which broadcast events as soon as a user interaction has been noticed. Whenever system-initiated interaction is necessary this can be done using three different interaction performatives: ASK presents an arbitrary question to the user, CHOICE asks questions with limited sets of answers, and VERIFY is used to query yes-no questions. The main advantage of this system is its ability to allow a natural interaction between the user and a distributed, dynamic and heterogeneous ensemble of devices and services.

3.3 Visual and Semantic Information Linking with Plant@Hand

On architectural level we use enterprise service bus (ESB) technologies to interface with heterogeneous software applications holding data required for our decision making process. These applications are coupled with the bus using application specific connectors. Then pre-configured information flows process the incoming data using our own context model. In this step a semantic filtering, which a) works implicitly through the structure of the context model and b) automatically identifies connections between structured and unstructured data using contextualized network graphs [5], generates additional links which are finally used to visually link the data on presentation layer. This way we give for example work sheets a specific location on the construction plan although there did not exist an explicit link before.

The final visualization in Plant@Hand follows known metaphors in engineering and production in order to reduce the cognitive effort of interpreting presented information. Key of our visual representation, as well as the users mental model of work to monitor, is the construction plan and the time schedule (see Figure 1). All other information is linked with both. Selecting a specific work task highlights the corresponding sections in both plans for example. On-site documentations are located exactly in the construction plan where they refer to.

3.4 Sensor Systems

To acquire context information regarding the users' current activities within our environment, we utilize various sensor systems. Among the most important sensor data is location information, i.e. positions and movements of persons which are then transformed to their symbolic equivalents. With this we can recognize e.g. the movement of the purchaser to another room corner.

Location data can be estimated by two different types of location systems: Passive and active location systems. With the latter require specific technical devices – mostly called tags – on part of the users to be able to estimate the location, the former work without them. However, they are not able to distinguish persons standing close to each other, which makes it impossible to determine the number of persons in the room solely based on those systems. To overcome this problem, we have developed a sensor data fusion system which models the sensor data of both systems together with the persons in the environment as a joint conditional density function and estimates the number of all persons together with their respective positions [6]. Thus, we have a unified location system able to deliver location estimations regardless of the underlying location systems, which may be active (e. g. Ubisense RTLS®) or passive ones (e. g. SensFloor®).

4 Approach

When building smart business applications we have to consider significant challenges:

- **Consistent user experience.** Business data and logic is typically located in heterogeneous information systems. Each comes with a different interaction strategy and metaphor, but smart business applications need to work with integrated information and to provide a consistent user experience.
- **Situation awareness.** Interacting with smart environments can be demanding for the user. Therefore, a situation aware assistance system should simplify the interaction. Instead of controlling available devices explicitly, the system should infer the current user goals and act proactively to support them. This allows the user to concentrate on his task instead.
- **Help users to understand.** Business applications provide complex functions for fulfilling work tasks. Smart business applications need to reduce learning efforts and cognitive load of users working with them.

Our use case application (see Figure 1) applies HCI research in order to improve the homogeneity and quality of working with data and information in business decision processes. We used a threefold approach consisting of situation modeling and intention recognition components, providing a consistent multi-touch interaction metaphor, and building the users' mental model of interaction accordingly.

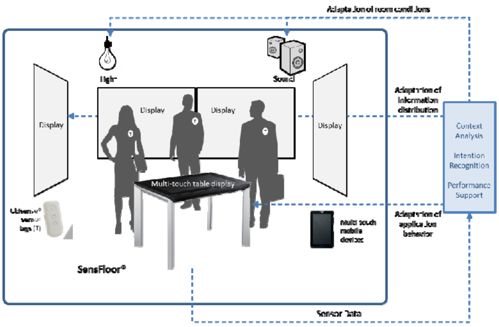

Fig. 2. Configuration of smart control room for providing smart business applications

We make use of the user's mental model in order to link work locations with work details and performance figures visually. The interaction takes place using well known gestures from working with paper documents but translated into their digital counterparts using multi-modal and multi-touch technologies. Our Plant@Hand application is finally embedded into a smart environment which provides us with sensors and opportunities to follow the user's movements, activities and their interaction as well as to estimate the user's intentions and thus next requirements with respect to the visual presentation of required data and information using multiple displays.

The next sections give a brief overview on our adoption of HCI research supporting the previously described decision making use case.

4.1 Situation Modeling and Intention Recognition in Smart Environments

Modern smart environments are equipped with various sensors and actuators. While actuators like projectors, canvases and audio equipment are used to present information, sensors are used to access the current state of devices and users. Intention recognition aims at inferring the current user goals from a sequence of sensor data [7]. For this, probabilistic models are constructed that are able to interpret noisy sensor data appropriately. To construct our intention recognition components, we employ causal behavior models, business process models, and sensor models [8]. These are automatically converted into a probabilistic model that can recognize the current abstract situation as well as the underlying user intentions and allows to support the users accordingly.

These causal models, used for translation to probabilistic models, are based on abstract action specifications. An action specification consists of a list of preconditions and effects to the environment. Together with a description of an initial environment state, such as "a critical issue appeared in the production process" and a description of a possible goal environment state, as “the issue has to be discussed and solved”, an probabilistic model that includes all possible action sequences leading from the initial state to the goal state can be generated. A mapping of possible environment states, such as “team-leader and supervisor are at the multi-touch table, the purchaser is moving to a private corner and there is an active Skype call”, to sensor data, including the current location of each user and the current Skype activity provides further information on the probability of possible environment states with given sensor information. Filtering techniques such as particle filter are then applied to determine current situation and thereby the probability of this initial state - goal state combination. A detailed description of this process is given in [9].

To recognize the intention of the team several combinations of initial states and goal states are selected and followed. Each hypothesis is then weighted as described and combined which results in a categorical probability distribution over possible goal states - the intention [10].

To reduce the number of possible intentions (combinations of initial and goal situations) sensor fusion techniques as described earlier are applied. If the prior probability that two persons are in the room is high for example, the system can skip all hypotheses that contain different numbers of acting agents. If later a third person (the purchaser in our case) enters these hypotheses have to be started again. The result of the intention recognition phase is a probability distribution of possible intentions, which is the base for further assistance decisions.

4.2 Multi-modal and Multi-touch Interaction with Information

In our use case the decision making takes place on the basis of structured and unstructured data from heterogeneous sources, e.g. ERP, MES, product data management systems (PDM), work sheets, or content management systems (CMS). The connection of data is seldom obvious as there are independent infrastructural software applications holding and providing it. Each application comes with a different user experience and requires a varying interaction style. This leads to additional cognitive loads required to switch between application contexts and usage metaphors [1].

In order to ensure a consistent user experience on multi-modal devices and to reduce the cognitive load of working with different applications required to draw conclusions and allow for decision making, we followed a two-folded approach:

- **Information integration.** We model, collect and interlink all required information prior to their further usage in our application. Starting with a context model of information as well as the process specific information requirements, we use *enterprise service bus technologies* to request the data from their native systems and aggregate it into our own model. So we bring together work plans from planning systems with construction data from product data management systems as well as shift reports from content management sources. The reports are semantically linked with work tasks which may have a location on the construction plan.
- **Context aware presentation and interaction metaphors.** If we want to reach a consistent user experience on multi-modal devices we have to provide a familiar way of presenting and interacting with information. So we combined the metaphors of industrial engineering (e.g. working with construction plans) with monitoring activities (e.g. controlling work sheets) in order to found a common experience. Our presentation and interaction modalities further address changing states of decision making, e.g. when we transfer certain information on public displays to allow an situational and adequate working.

The information integration takes place on data level. On top of this we built with the Plant@Hand application a consistent multi-touch user experience on multi-modal devices. In our stationary control room scenario we transferred the metaphor of using paper documents on a table into an implementation on a multi-touch table interface where this familiar way of interacting can be found again. Touch gestures or digital pen annotation known from the classical paper metaphor are used to interact with construction plans, time figures as well as text, image, or video documents. The same interaction is made available through Plant@Hand in a mobile decision making scenario, where we transferred the visual and interactive part of our application onto mobile devices, e.g. smartphones or tablets.

The intention recognition allows us finally to adapt the visual representation and smart environment accordingly. Dependent on a recognized situation during the team discussion (monitoring, briefing, planning, problem solving) , room conditions (light, acoustics, projection modes, display usage) and application behavior (presentation mode, information distribution) are modified (see Figure 1).

4.3 Shaping the User's Mental Model of Interaction

Efficient interaction with intelligent environments is a challenge, due to the number of possible options and offered functionality. Novice users are often overwhelmed while entering our lab, a smart meeting room equipped with various sensors and actuators, which served as a hardware platform for this work. A consistent user experience is needed to simplify the interaction. In [1], some design principles have been identified to reduce learning efforts and cognitive load, in particular observability, transparency, and tangibility. Those principles were applied while developing the system described here.

In smart environments both user and application need to understand the counterpart's actions and intentions to align own actions accordingly. Analyzing the user's intentions has been discussed above, here we concentrate on how to shape the user's mental model of the environment. A sufficiently correct mental model is necessary to understand and predict the behavior of the system. Usually, the behavior of such systems is described in manuals provided by the manufacturers. But this is not possible in our setting, because the environment, i.e., the system itself, is dynamic – devices, services and users can enter and leave the ensemble at any time. Therefore, the overall behavior can neither be fixed at the time of the installation, nor can it be described in a manual. Nonetheless, the behavior must be understandable for the user. We believe that a solution to this problem is the system's ability to explain its behavior and the ability to teach the user. As a result, the user's mental model is shaped and updated by the environment itself. The self-explanation capability enables the users to learn more about the system without referring to external sources [11]. A realized controller with self-explanation capabilities is described in [12].

Most people tend to understand technical systems as “simple” stimulus-response systems (e.g. normal software applications, ticket machines). To achieve a goal, they need to perform an explicit triggering action, e.g., to switch on the projector, they need to use the corresponding interface. This mental model influences their expectations and interaction. Earlier work in psychology [13, 14] showed this impact on learning to operate new devices, for example. Instead of this automata model, we would like the user to perceive the environment as being controlled by a proactive intelligent assistant. To achieve this goal, the environment should be able to answer questions with respect to its current state and it should be able to teach the user to use it efficiently. For example, if the system recognizes that the user performs a number of actions to prepare a presentation on a public display as needed in the simple automata model, a message could be generated describing how to use the room more efficiently. Here we can adopt scripting approaches from computer supported cooperative work research (CSCW). There scripting is a well-known approach to structure the learning of new methods. It is also used to guide as well as organize cooperative interactions between people [15]. Unfortunately, and to the best of the authors’ knowledge, research in this direction is just about to evolve but no real solutions exist so far.

5 Results and Evaluation

Evaluating the overall performance increase of such a complex system is hard or even impossible without long term studies pursued on-site. Instead, we evaluate different aspects of the system independently.

The evaluation of the situation recognition was done in [9] based on a similar problem from the meeting domain. The location data of three persons was recorded during several different meetings and later used as input for the inference engine. The evaluation illustrated that the described approach is (1) competitive with probabilistic models built from training data, (2) able to provide activity recognition with 90% accuracy and (3) reusable in different applications within the same domain. The intention recognition based on these causal models was introduced in [10]. A live demonstration showed that recognizing the intention based on activity recognition with causal models is valid and provides accurate results.

Additionally we brought our Plant@Hand business application on-site into a manufacturing company for cooling devices and evaluated there with staff under field conditions. The results show a positive effect of picking-up the user's mental model with our chosen visualization and interaction styles. After a short learning phase of basic touch gestures (move, resize) the assembly team worked autonomously with the application. With respect to efficiency the main benefit was seen in the information integration avoiding time consuming search in paper documents or opening different applications to collect the required data. The evaluation also showed a fast adoption of our provided information linking methods. After working a while with the application the team was requesting a construction plan manipulation on design level which shows a seamless transition from monitoring and controlling towards re-planning.

6 Conclusion and Future Work

Research in HCI progresses continuously towards a fast adoption of novel device generations which vice versa enable new interactions with the user. With addressing a specific use case from manufacturing industries we focused on bringing together different HCI approaches to improve the efficiency and quality of assistance in business decision making. By realizing the system within our smart lab environment, we showed the principal applicability of this approach. First evaluation results, especially with respect to intention recognition and the multi-modal interaction with information, demonstrate not only a general usability but also a growing user acceptance. Nevertheless, there is more research needed to provide proactive and self-explanatory abilities within such smart environments. Here we will address with our future work the role of mental models on the contextualization of visualization and interaction technologies as well as on the learnability of autonomously working systems.

Further work and research is also addressing the transfer of our results into other usage scenarios. Here we aim to support mobile on-ship monitoring as well as local control rooms to track and manage fishing fleets and their by-catch.

Acknowledgment. Sebastian Bader is supported by the DFG within the GRK1424 MuSAMA.

References

1. Aehnelt, M., Peter, C., Müsebeck, P.: A Discussion of Using Mental Models in Assistive Environments. In: Proc. of the 5th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA 2012). The 5th Workshop on Affect and Behaviour Related Assistance (ABRA), Crete, Greece, June 6-8 (2012)
2. Bader, S., Ruscher, G., Kirste, T.: Multimodal Interaction in Dynamic and Heterogeneous Smart Environments. In: Proc. of the 1st International Workshop on Language Technology in Pervasive Computing (LTPC) at the PerCom, UK (2012)
3. Bader, S., Kirste, T.: An Overview of the Helferlein-System. Rostock, Germany, Technical report CS-03-12 under (2012) ISSN 0944-5900
4. Heider, T., Kirste, T.: Supporting goal-based interaction with dynamic intelligent environments. In: Proc. of ECAI, pp. 596–600 (2002)
5. Ceglowski, M., Coburn, A., Cuadrado, J.: Semantic search of unstructured data using contextual network graphs. National Institute for Technology and Liberal Education 10 (2003)
6. Ruscher, G., Kirste, T.: Estimating User Number and Location from a Multi-Modal Localization Sensor Setup. In: Proceedings of the UPINLBS 2012, Helsinki, Finland (2012)
7. Krüger, F., Ruscher, G., Bader, S., Kirste, T.: A Context-Aware proactive Controller for Smart Environments. I-COM 10, 41–48 (2011)
8. Yordanova, K., Krüger, F., Kirste, T.: Strategies for Modelling Human Behaviour for Activity Recognition with Precondition-Effect Rules. In: Proc. of the 35th German Conference on Artificial Intelligence, Saarbrücken, Germany (2012)
9. Krüger, F., Yordanova, K., Hein, A., Kirste, T.: Plan Synthesis for Probabilistic Activity Recognition. In: Proceedings of the ICAART 2013, Barcelona, Spain (2013)
10. Krüger, F., Yordanova, K., Kirste, T.: Tool support for probabilistic intention recognition using plan synthesis. In: Paternò, F., de Ruyter, B., Markopoulos, P., Santoro, C., van Loenen, E., Luyten, K. (eds.) AmI 2012. LNCS, vol. 7683, pp. 439–444. Springer, Heidelberg (2012)
11. Bader, S.: Probably Asked Questions: Intelligibility through Question Generation. In: Mandl, S., Ludwig, B., Michahelles, F. (eds.) Proc. of the 2nd Workshop on Context Aware Intelligent Assistance, Berlin, Germany (2011)
12. Bader, S.: Explaining the Reactions of a Smart Environment. In: Proc. of Exact 2012, Montpellier, France (2012)
13. Johnson-Laird, P.N.: Mental models: towards a cognitive science of language, inference, and consciousness. Harvard University Press (1983)
14. Kieras, D.E., Bovair, S.: The role of a mental model in learning to operate a device. Cognitive Science 8, 255–273 (1984)
15. Aehnelt, M., Hambach, S., Müsebeck, P., Musielak, M., Hoog, R., de, K.J., Lindstaedt, S.: Context and scripts: supporting interactive work-integrated learning. In: O'Malley, C., Suthers, D., Reimann, P., Dimitracopoulou, A. (eds.) CSCL 2009: In Proc. of the 9th International Conference on Computer Supported Collaborative Learning, vol. 1, pp. 144–146. International Society of the Learning Sciences (2009)