REAL: Situated Dialogues in Instrumented Environments *

Christoph Stahl, Jörg Baus, Antonio Krüger, Dominik Heckmann, Rainer Wasinger, Michael Schneider
Saarland University
66123 Saarbrücken
Germany
stahl@cs.uni-sb.de

ABSTRACT

We give a survey of the research project REAL, where we investigate how a system can proactively assist its user in solving different tasks in an instrumented environment by sensing implicit interaction and utilising distributed presentation media. First we introduce the architecture of our instrumented environment, which uses a blackboard to coordinate the components of the environment, such as the sensing and positioning services and interaction devices. A ubiquitous user model provides contextual information on the users characteristics, actions and locations. The user may access and control their profile via a web interface. In the following, we present two mobile applications to employ the environmental support for situated dialogues, a shopping assistant and a pedestrian navigation system. Both applications allow for multi-modal interaction through a combination of speech, gesture and sensed actions such as motion.

Keywords

Multi-modal speech and gestural interface, transparent user modelling, instrumented environment, pedestrian navigation

1. INTRODUCTION

The project REAL is concerned with the main question: How can a system assist its user in solving different tasks in an instrumented environment? Such environments consist of distributed computational power, presentation media and sensors as described in section 2, and also entail the observation and recognition of implicit user interactions in the environment. This offers the possibility to infer about a user's plan(s) and intentions, and to proactively assist in solving their task.

We focus our interest on two particular tasks in an airport scenario, shopping and navigation. In the shopping scenario, we explore how to assist the user in achieving their goal of the best possible buying decision within a given limited time. We employ an RFID-technology based infrastructure of readers and labeled products to sense implicit user interaction, such as picking up a product from the shelf or putting it into the shopping cart. We can imagine a variety of mobile and stationary devices to be used for information presentation and dialogues with the user. Some user might prefer their own PDA or smartphone, others might use a shopping cart equipped with a touchscreen. It is also desirable to involve large public displays for the presentation of rich media content, which otherwise are used to display advertisements. Our current assistant is introduced in section 4. In order to enter the shop and to pick up a certain product, the user pursues certain navigational goals. Therefore we are integrating our previously developed stand-alone pedestrian navigation system into the instrumented environment, which provides for routing and positioning services. Besides the navigational aid, the system also offers an exploration mode, which allows the user to query information on points of interest within a three-dimensional navigational map. The user may formulate their request using combined speech and stylus gestures, as described in section 5.

2. ARCHITECTURE

The architecture of the intelligent instrumented environment is shown in Figure 1. The layout vertically arranges the components in layers, representing the flow of information from the physical devices upwards to the logical services. The components utilise a common blackboard for their communication, which they use to read requests and to write response messages. The diagram also distinguishes between personal and public devices, which are ordered from left to right. In the positioning-layer, we adopt infrared and radio-frequency technology in two complementary approaches. As shown on the left, we mark the user or any mobile object with a visual or radio-frequency identification tag. We use the knowledge about the location of the receiving camera or antenna to estimate the object's position. We also deploy a public beacon infrastructure in the environment. The beacons are used to provide a position identification signal for the pedestrian navigation component, which will be explained in further detail later in this section. The interaction-layer comprises the physical devices used for interaction. The left hand side of the diagram shows the user's own personal digital assistant and the shopping cart equipped with a tablet PC, which is temporarily associated with a single user. They provide the user interface for the navigation and shopping assistants. Other components of the assistant, such as sensing technology, implemented services and knowledge, are clustered vertically above them. Next to the personal devices, the instrumented shelf is shown. Even though it is inherently public, we currently only allow a single user to pick up products at any one time (to avoid misinterpretations). The smart-door-displays provide a public interface

*REAL is a Project of the Collaborative Research Program 378 "Ressource-Adaptive Cognitive Processes"

ITI Workshop, Part of AVI 2004 (Advanced Visual Interfaces) 2004 Gallipoli (Lecce), Italy

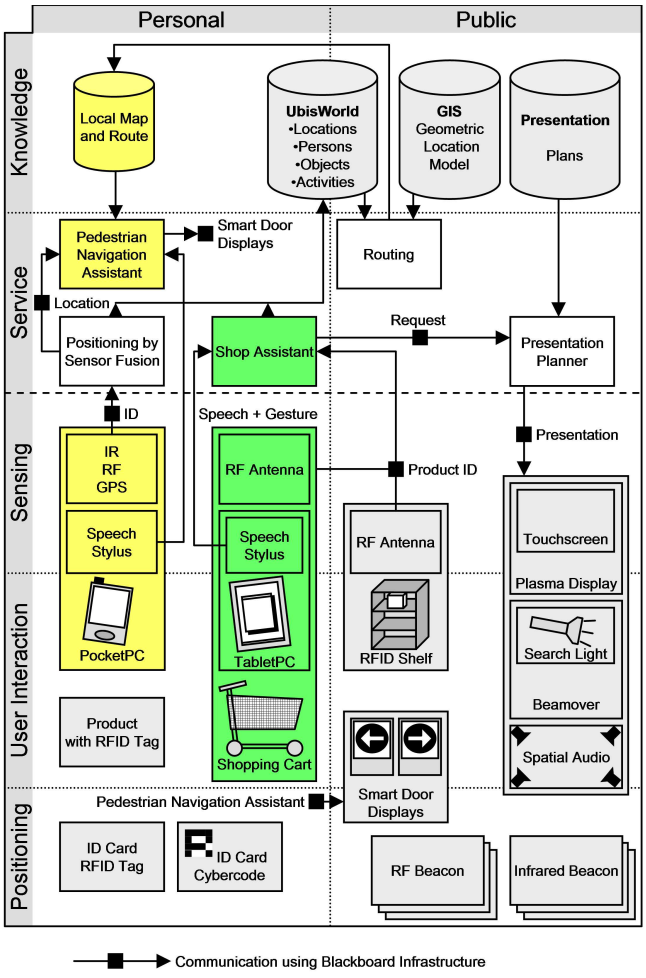

Figure 1: Overall system architecture

to enter messages for the room inhabitants. They also present personal navigation aid to users in their vicinity. On the right hand side, the public interaction devices provided by the environment are shown, such as various public displays and a loudspeaker system. On the sensing layer, the diagram depicts the technologies which are available on a device to sense the user interaction. The Pocket PC and Tablet PC both provide a microphone for speech input and a touch-screen for stylus interaction. The PDA has a built-in infrared sensor to receive beacon ID's and optionally an RF reader to scan for RF beacons. Additionally, it is equipped with a bluetooth GPS receiver for outdoor use. The collected positioning signals are sent via the blackboard to the positioning service. The cart and the shelf are instrumented with an RF receiver/antenna. They continually scan for RFID-tagged products and evaluate changes as pick-up or put-down gestures. The service layer draws the borderline between physical and logical components. It provides the services of the environment, which implement the necessary functionality for the navigation and shopping assistance applications. The services will be explained in the application sections 4 and 5. The top layer comprises knowledge about location, user, activity and presentation planning. The location model consists of a geometric and symbolic model, the latter is provided together with the user model by the ubiquitous world model called UbisWorld.

3. UBIQUITOUS USER MODELLING

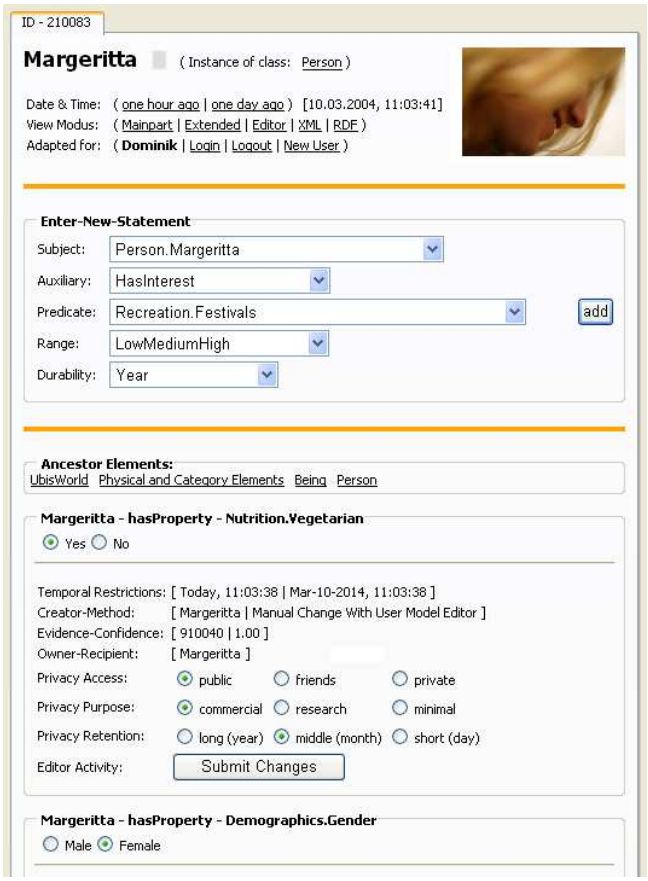

Figure 2: Transparent web interface to the user model

Ubiquitous computing in an instrumented environment as described above poses challenges and new applications to the field of user modelling. A specialised markup language USERML and the appropriate query language USERQL have been developed to allow for the exchange and distributed storage of partial user models and context models (Heckmann & Krüger, 2003). The underlying model of so called "Situational Statements" is based on ideas of the semantic web with RDF/RDFS[1], which means that complex descriptions about *resources* can be uniformly defined. Such resources are identified with qualified Uniform Resource Identifers (URI)[2] with the slight difference that they have an optional fragment identifier with an additional part separated by a dot. Situational statements are introduced in (Heckmann, 2003) as an integrating data structure for user modelling, context-awareness and resource-adaptive computing.

RDF descriptions, which together with the reification mechanism form a powerful knowledge representation framework, are represented by triples. Situational statements can however be seen as *Extended Triples*, that extend the basic subject-predicate-object triple with temporal- and spatial restrictions as well as meta-data about ownership, and privacy, evidence and confidence. A whole collection of situational statements is called a *Situation*. As counterpart to the exchange language USERML, which forms an or-

[1]http://www.w3c.org/RDF
[2]http://www.w3.org/Addressing/

dinary XML application, the domain ontology USERMODEL ONTOLOGY has been developed to introduce concepts of the user modelling research area as referable resources in the semantic web languages DAML[3] and OWL. One advantage of this modularised approach in which the ontology and the representational formalisms are separated, is that everybody could define their own ontology while using the same representation and tools. The ontology defined for our instrumented environment is called UBISWORLD ONTOLOGY and apart from the user model covers all elements from physical objects, locations, times, their properties and features, activity and inference elements (Wasinger et al., 2003).

From the user perspective, it is crucial to have insight into the user model and to control privacy levels. Our user model server provides a comfortable web-based interface to transparently access the properties of any modeled object within the environment. Figure 2 shows the properties of the user *Margeritta* as an example. In the upper section, the web form allows the user to enter new properties as statements, by using the UBISWORLD ONTOLOGY. In our example, we are going to express that Margeritta has a high interest in festivals for recreation. Besides, the interface gives a listing of her modeled properites. For example, the first property says that she is a vegetarian. This statement about her nutrition habits is set by herself to be public and may be used by commercial services, like the shopping assistant presented in the following section. If the statement would have been the result of an automated inference, she could see the origin of creation and the confidence value of the assumption. If the inference were wrong, Margeritta could manually override the assumption.

4. SMART SHOPPING ASSISTANT

In contrast to explicit human-computer-interaction, where a user controls an application's behaviour through buttons, menus, dialogs, or some kind of written or spoken command language, one of our first applications developed for the instrumented environment (see Section 2) uses implicit interactions and is driven by the user's actions in the environment. The prototype of an intelligent, adaptive shopping assistant, named SMART SHOPPING ASSISTANT (SSA), offers value-added services in a shopping scenario (see also Schneider, 2003). The assistant uses Radio Frequency IDentification (RFID) sensors and plan recognition techniques to transparently observe a shopper's actions. From the observed actions the system infers the shopper's goals. Using this information a proactive mobile assistant mounted on a shopping cart offers support during shopping. The type of value-added services offered, e.g. while buying groceries, range from simple product comparisons, and analysis of goods in the shopping cart for cross-selling recommendations, to the suggestions of recipes. Another speciality of the system lies in its ability to make use of a user model in order to adapt value-added services to special preferences (e.g. nutrition habits) and needs of the individual customer.

During shopping, relevant user actions for example include moving around in the store, looking at items of interest or advertising displays, or physically interacting with products. Relevant context for example includes a user model as provided by the aforementioned USERMODELSERVER (see section 3), a location model, and the products available in the store or involved in a user's action.

The SSA comprises a central server, a modified shopping cart or basket, and instrumented shelves. Both, the shopping basket and the shelf, have been equipped with RFID readers as shown in the first image of figure 3, to allow for the recognition of products tagged with RFID transponders. In addition to the basket, a PDA is used as the primary user interface. Alternatively to the basket, we have a shopping cart with integrated RFID antenna and a tablet PC mounted on the cart's handrail, hosting the shopping assistant, controlling the RFID reader, and connecting the cart with the central server via wireless network. With this setup, user actions like taking a product and putting it down can be recognised by repeatedly polling the transponder's IDs in the antenna field of each shelf and shopping cart. These observations are fed into the shopping assistant application, which is provided as a service by the intelligent environment. The application reacts to these observations and the context provided by the user, and offers value-added services like product comparisons or cross-selling recommendations. An example is shown in figure 3. The information is sent to the presentation planning service, which selects an appropriate display and generates a resource adaptive presentation.

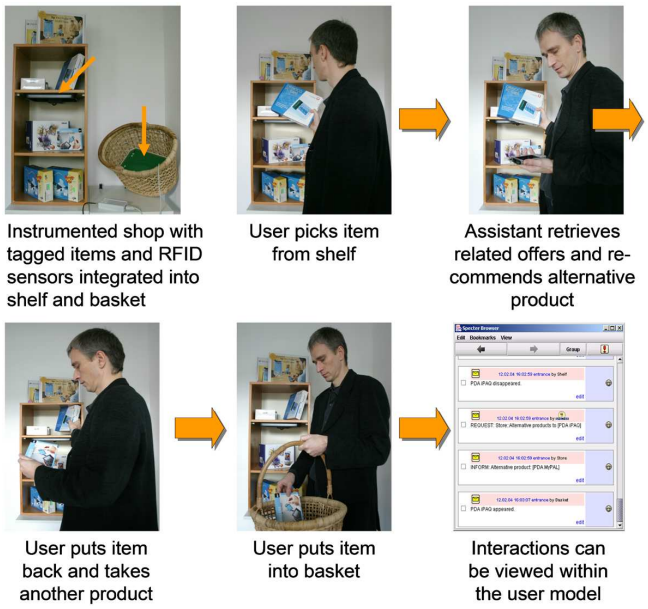

Figure 3: User interactions serve as input for the assistant

5. PEDESTRIAN NAVIGATION ASSISTANT

Being at the right place at the right time is a essential precondition for any user interaction within the real world. Thus for any given user task, it is most likely that a navigational sub-goal exists. Being aware of this, we have spent considerable research effort in the development of indoor and outdoor pedestrian navigation systems, as published in (Baus, Krüger, & Wahlster, 2002), and recently the BMW PERSONAL NAVIGATOR (Krüger et al., 2004). Whereas the predecessor systems have been designed as independent devices, the new navigational assistant utilises the positioning services and infrastructure provided by the instrumented environment.

The mobile device sends all positioning data received by GPS and beacons to this service, which fuses the different sensor information (like IR and RF beacon-IDs and geo-coordinates by GPS) and matches them with spatial knowledge of the environment. The positioning service returns a symbolic location identifier as well as geometric coordinates for map visualisation and the generation of situated navigational aid. Optionally, the mobile user may specify their user profile, in order to update the location information in the

[3] http://www.daml.org/ontologies/444

ubiquitous user model UbisWorld. This is our framework to enable adaptive user-interfaces for situated interaction, and location based services in general (see section 3). The user interface of our pedestrian navigation assistant is based on a three-dimensional interactive map (see Figure 4 A). The use of 3D VRML landmark models allows for different perspectives, such as a plan view and a tethered view, as well as pan and zoom capabilities. The assistant supports two different modes of operation: *navigation* and *exploration*.

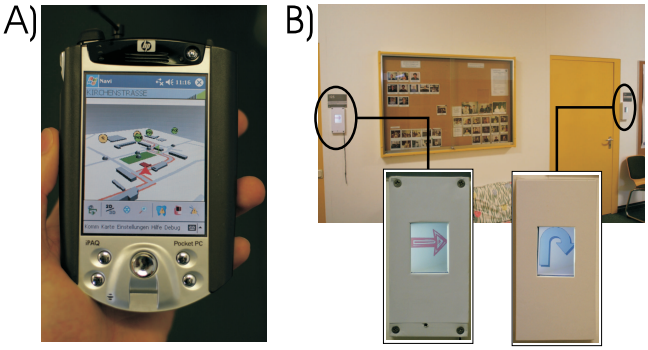

Figure 4: Navigation aid is provided by the mobile navigation assistant A) and (stationary) smart door displays B)

The prior allows the user to plan a route to a certain destination, using the environments routing service. These routes can be either indoor or outdoor. Upon downloading the route, the user can be directed along the route through the use of graphical, acoustic and spoken navigational aids. In this mode, the system responds primarily to the sensed change of location and requires no additional user input.

Aside from the navigation mode, the user can at any time switch into an exploration mode. In this mode, they can leave their planned route and explore their surrounding environment, querying what it is that they see. Often these user-queries take the form of combined speech and gesture, in which the user can speak and point simultaneously at objects on the PDA display in front of them. Common user-requests (currently in German) take the form of: *"What is that?"* and *"describe this landmark to me"* (for more details see Wasinger, Stahl, & Krüger, 2003).

Other functions of the navigation system include a memo feature that stores notes (text / graphics / speech) to the user's localised position on the map, and the ability to re-listen to position-sensitive information regarding distance to nearby landmarks.

Like its predecessor systems, the user interface is designed to adapt to the user's current situation and resources (see Kray, 2003; Wasinger et al., 2003). Although the combination of speech and gesture offers the highest grade of flexibility to the user, in certain situations either modality may not be appropriate. Thus the system may be solely used by a traditional graphical interface, or solely without the need of vision. This is a new and challenging situation, which applies for freehand operation, for devices without display, or visually impaired users. In our system, we implemented voice control for all menu items and we use speech synthesis and 3D spatial audio cues to convey information about the distance and orientation of nearby landmarks. A first user study with a blind student has been promising. Although the GPS positioning accuracy was found to be insufficient for navigational aid, the user did appreciate the information on streets and landmarks.

Indoors, the navigation system utilises the smart door displays to present visual and acoustic guidance information on location (see Figure 4B). The devices are based on PocketPC PDAs that are wall-mounted next to the regular door displays. They have been primarily designed to communicate messages between the absent room inhabitant and visitors. A second role is to present arbitrary SMIL encoded multimedia content on request via the blackboard architecture. The mobile navigation system's map contains activation areas for the smart door displays along the way. Upon entering such an area, the assistant assigns the display to show route directions. Since the displays are rather small, an individual sound is used to draw the user's attention to them.

6. RELATED WORK

The CONTEXT TOOLKIT (Salber, Dey, & Abowd, 1999) aims to support the development of context-aware applications within computationally-enhanced ubiquitous computing environments. Context-aware information services adapt to any information that can be used to characterise the situation of a user, such as their location and objects and persons in the vicinity.

The Stanford Interactive Workspace IROOM (Fox, Johanson, Hanrahan, & Winograd, 2000; Schreyer, Hartmann, Fischer, & Kunz, 2002) is a collection of linked software and hardware that allows users to interact with their application suite on three large smartboard displays. The iRoom architecture provides the EVENTHEAP tuplespace for the coordination of its components. In comparison with the event messaging concept, the persistence of information within the tuplespace leads to several advantages, such as dynamic coordination, failure tolerance and anonymous communication.

The Ambiente project at the Fraunhofer IPSI (Institut Integrierte Publikations- und Informationssysteme), persues research on interactive communications- and collaboration landscapes. Several so-called Roomware components and artefacts have been developed . Their BEACH software (Tandler, 2003) provides a model for the creation and handling of hypermedia data and for collaboration.

The Hermes project (Cheverst, Dix, Fitton, & Rouncefield, 2003) has equipped two corridors of the Department of Computing at Lancaster University with interactive door displays based on handheld devices. On the borderline between private and public space, they provide simple asynchronous graphical and textual messaging services between office occupants and anyone passing by their office.

In the Smartkom (Wahlster, 2003) framework, a mobile handheld assistant was developed to process multi-modal input and output (speech and gesture) during a navigational task. The presented information combined maps, natural language and a life-like character.

In 2003, the METRO Group opened its FUTURESTORE[4] in Rheinberg, Germany, to the public. RFID technology is used to track the location of each individual product in order to develop new processes in inventory management. The store also offers new information possibilities for the customer, such as a personal shopping assistant, information terminals and advertising displays, all based on the RFID tags.

7. OPEN ISSUES

As stated by this survey, we have until now solely considered interactions with a single user in our instrumented environment. Multi user interaction in instrumented environments raises many new research questions (Kray, Wasinger, & Kortuem, 2004). There is an increased interest in how to confer our methods, techniques and concepts developed so far, to allow for the interactions of mul-

[4]http://www.future-store.org/

tiple users or user groups at the same time in our instrumented environment. Some concepts have already been developed and published in (Kruppa, 2004; Rocchi, Stock, Zancanaro, Kruppa, & Krüger, 2004).

References

Baus, J., Krüger, A., & Wahlster, W. (2002). A Resource-Adaptive Mobile Navigation System. In *IUI2002: International Conference on Intelligent User Interfaces* (pp. 15–22). New York: ACM Press.

Cheverst, K., Dix, A., Fitton, D., & Rouncefield, M. (2003). Out to lunch: Exploring the sharing of personal context through office door displays. In *Proceedings of International Conference of the Australian Computer-Human Interaction Special Interest Group (OzCHI'03)* (pp. 74–83).

Fox, A., Johanson, B., Hanrahan, P., & Winograd, T. (2000). Integrating information appliances into an interactive workspace. *IEEE Computer Graphics and Applications*, *20*(3), 54–65.

Heckmann, D. (2003). Introducing situational statements as an integrating data structure for user modeling, context-awareness and resource-adaptive computing. In A. Hoto & G. Stumme (Eds.), *LLWA Lehren - Lernen - Wissen - Adaptivität (ABIS2003)* (p. 283-286). Karlsruhe, Germany.

Heckmann, D., & Krüger, A. (2003). A user modeling markup language (UserML) for ubiquitous computing. In P. Brusilovsky, A. Corbett, & F. de Rosis (Eds.), *User Modeling: Proceedings of the Ninth International Conference, UM2003* (pp. 393–397). Johnstown, PA, USA: Springer.

Kray, C. (2003). *Situated Interaction on Spatial Topics.* Akademische Verlagsgesellschaft Aka GmbH.

Kray, C., Wasinger, R., & Kortuem, G. (2004). Concepts and issues in interfaces for multiple users and multiple devices. In *Workshop on Multi-User and Ubiquitous User Interfaces (MU3I) at IUI/CADUI* (pp. 7–12).

Krüger, A., Butz, A., Müller, C., Stahl, C., Wasinger, R., Steinberg, K., & Dirschl, A. (2004). The Connected User Interface: Realizing a Ppersonal Situated Navigation Service. In *Proceedings of International Conference on Intelligent User Inferfaces.* ACM Press.

Kruppa, M. (2004). The better remote control - multiuser interaction with public displays. In *MU3I workshop at International Conference on Intelligent User Interfaces.* ACM Press.

Rocchi, C., Stock, O., Zancanaro, M., Kruppa, M., & Krüger, A. (2004). The museum visit: Generating seamless personalized presentations on multiple devices. In *Proceedings of International Conference on Intelligent User Inferfaces.* ACM Press.

Salber, D., Dey, A. K., & Abowd, G. D. (1999). The context toolkit: Aiding the development of context-enabled applications. In *CHI* (p. 434-441).

Schneider, M. (2003). A Smart Shopping Assistant utilizing Adaptive Plan Recognition. In A. Hoto & G. Stumme (Eds.), *Lehren - Lernen - Wissen - Adaptivität (LLWA04)* (p. 331-334).

Schreyer, M., Hartmann, T., Fischer, M., & Kunz, J. (2002). *iRoom XT design and use* (technical report No. TR144). Stanford CIFE.

Tandler, P. (2003). the BEACH application model and software framework for synchronous collaboration in ubiquitous computing environments. *Journal of Systems and Software, Special Edition on Application Models and Programming Tools for Ubiquitous Computing*.

Wahlster, W. (2003). Towards Symmetric Multimodality: Fusion and Fission of Speech, Gesture, and Facial Expression. In A. Günter, R. Kruse, & B. Neumann (Eds.), *Proceedings of the 26th German Conference on Artificial Intelligence* (pp. 1–18).

Wasinger, R., Olivia, D., Heckmann, D., Braun, B., Brandherm, B., & Stahl, C. (2003). Adapting spoken and visual output for a pedestrian navigation system, based on given situational statements. In A. Hoto & G. Stumme (Eds.), *LLWA Lehren - Lernen - Wissen - Adaptivität (ABIS2003)* (p. 343-346). Karlsruhe, Germany.

Wasinger, R., Stahl, C., & Krüger, A. (2003). Robust speech interaction in a mobile environment. In *Proccedings of the 8th European Conference on Speech, Communication, and Technology (Eurospeech '03)* (pp. 1049–1052).