Enabling Natural Interaction by Approaching Objects

Andreas Lorenz, Andreas Zimmermann and Markus Eisenhauer
Fraunhofer FIT
Schloss Birlinghoven
53754, Sankt Augustin, Germany
{andreas.lorenz, andreas.zimmermann, markus.eisenhauer}@fit.fraunhofer.de

Abstract

This paper introduces the research issues to make a major step towards natural human interaction with mobile multimodal systems. Thereby, we aim at combining explicit and implicit interaction to blended multimodal interaction. Natural interaction and the requirement to anticipate a mix of purposive behaviour and uncertainty requires rendering information in a way that is compatible with and aware of the situational context. In two application examples we illustrate our approaches of interaction by movement. As the basis for adaptive behaviour, this paper describes the conclusions drawn from tracking the mobile user in two environments: a museum and a maintenance setting. It also illuminates different aspects of this technology, such as accuracy and costs, and illustrates the results of our project using the specific technology.

1 Introduction

Technology available today allows realizing a wealth of services for mobile people. Such services may support mobility as such, e.g. supply information about transportation or other facilities needed by travelling or commuting people. Or the services may simply supply any sort of information and service that needed while moving, e.g. shopping, tourism or task-related content. In this paper, we refer to such services as mobile guides.

The success of future mobile systems and applications will be highly affected by user-friendly and intuitive humanlike interaction procedures. The raising concern on accessibility and usability are pushing research into exploring and employing different modalities of information input and output. For example, if mobile users are moving around, they cannot easily devote all of their visual attention for interacting with a graphical interface [Brewster, 2002]. Conversational multimodal interaction will be a paramount requirement to enable seamless ubiquitous computing and therefore for the acceptance of third generation mobile services by the market. Most of the research and design efforts were invested so far in Graphical User Interfaces; nevertheless, experiments and research in multimodal environments have proved the potential efficiency of combining different communication modes.

According to Oviatt's myths of multimodal interaction [Oviatt, 1999], speech and pointing as the dominant multimodal integration pattern has been widely overrated and other modalities have been culpably neglected so far. Speech as the primary input mode doesn't work (e.g. in noisy environments) and other modalities might be preferred under such circumstances. In order to achieve natural interaction, such changing environments have to be taken into account and the benefits of other modalities have to be addressed.

In this paper we explore a user-centred approach in the design of systems supporting blended multimodal interaction. The goal is to make a major step towards natural human interaction with multimodal systems, by endowing these systems with humanlike cognitive capabilities that are necessary to interact with real users, especially in applications for which extensive user training is not feasible. We introduce a new approach to interaction that exploits the natural spatial motion of actors and objects to each other. In order to exploit spatial relations the motion of users and objects must be tracked. The main information source for the system obviously is a tracking system. This paper starts with an overview on the localization and tracking solutions we are using. In connection with a collection of principles leading us towards natural interaction, we will describe two projects making use of the user's body as an interface. First, we describe our experience from the LISTEN-project relating on high-accurate tracking-information. Then we will describe our current project MICA combining different technologies.

2 Localization and Tracking

One major parameter for present context-adaptive systems is the location of objects as an important dimension for the determination of the context of the object. A location is defined as a point within a defined environment. In addition, the location can be *relative* related to other objects of this environment or *absolute* related to a fixed coordinate system.

The access to the locations of relevant objects enables context-adaptive systems to draw conclusions about relations of objects or people to each other in a physical environment, e.g. by which objects a specific entity is encircled , what objects are in eyesight or in an acoustic range of a user, or where entities do move to. These interrelations of objects in an environment can be used by the system to conclude what information might be needed. The system can react to the context of the object, e.g. to provide a user with information about close objects or to guide a user to a requested location.

The accuracy and latency of the determination of the user's location depends on the requirements of the application. The more immersive the application environment needs to be, the more accurate, real-time and distinctive the provision of the location has to be.

2.1 Clarification of Terms

Most of the systems make use of a mixture of localization and tracking systems providing location information and do not take an explicit distinction between them. The terms localization system and tracking system are being used synonymously since normally it is the specification for real-time location-tracking systems. For this reason we will define the terms *localization* and *tracking*, before we describe several approaches for acquiring this information.

Localization: A location system determines its position and orientation within the environment by measuring distances or angles to known reference points. The system translates the relative position into absolute coordinates. This process, usually referred to as positioning, makes use of given external information and use the location information internally.

Tracking: If the system surveys single persons/objects and informs them about their position but the person/object does not participate actively in the process of localization, the term tracking is being used [Hightower and Borriello, 2001]. The surveyed person/object cannot influence the tracking process directly.

The major difference between both approaches lies in the handling of the identity of the localized object or person. In contrast to pure localization systems which analyze the information of the position locally and where the position or identity is not necessarily forwarded to a central system, tracking systems have a central unit that has all information of the position and identity of the persons/objects.

2.2 Localization and Tracking Systems

This section provides an overview on current indoor localization and tracking systems which are working with. In addition, we give assessments concerning their precision and efficiency.

Optical Tracking

Optical systems are equipped with cameras observing the area. Basically, we found two methods for tracking: Image recognition and retro-reflection.

With image recognition, positions and movements are computed from analyzing the scenery or superposed scenes. This approach falls back on complex algorithms and requires high-quality imaging, which limits the application fields and boosts costs. The method extracts features from the images like edges, colour gradients and contrasts to identify the object. Once an object is identified, useful information about this object can be delivered and objects in the near range can be used proactively for recommendations to the user. Furthermore, the objects in the focus of the camera can be used to determine the location of the user, e.g. the current room. To reduce complexity, the search-space of possible objects can be narrowed to those objects in the current range of the user. An example of an image recognition system has been developed and validated in [Nuamah, 2004].

The camera based system like [A.R.Tracking, 2005] uses retro-reflecting (passive), mostly spherical markers for marking the subject or object to be tracked. These markers are recognized by the tracking cameras. The body to be tracked (e.g. a human body or an object) is equipped with markers that are often covered by retro-reflective surfaces (see Figure 1). Tracking cameras are image sensors working in the near infrared light spectrum and recognize these markers using an infrared light flash (not visible for the human eye), which illuminates the measurement volume periodically. A central signal processing unit collects the data accumulated by receivers and calculates the absolute position and the orientation of each tracked object.

For a proper determination of the position and the orientation at least four receivers with direct line-of-sight contact to the transmitters are required. For the observation of large or complex rooms, as well as for an improved accuracy and reliability the system is scalable and more receivers can be deployed. Out of these receivers the system automatically selects the most favourable one for position/orientation calculation. Non-line-of-sight links as well as interfering multi-path signals will be discarded by plausibility checks concerning the level of the RF-signal received (signal quality) and the position calculated.

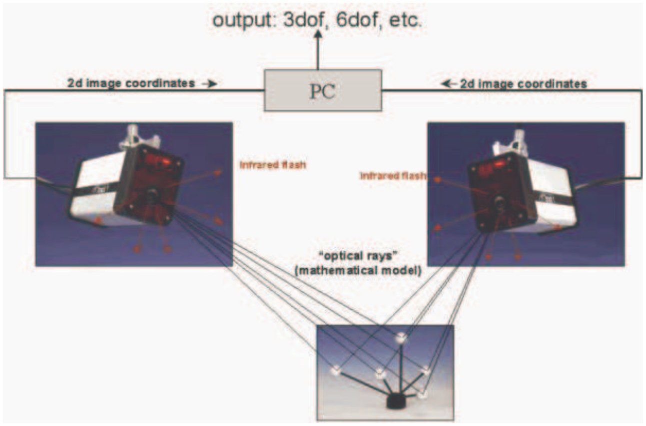

Figure 1 Tracking Principle of the ARTracking-System

With a minimal setup the tracking system is able to measure the positions of 8 users simultaneously with an absolute accuracy of the head position of about 10 cm. The minimum granularity of the orientation angle comes to 5 degrees. Therefore, the system requires precisely calibration of the cameras.

Wireless-LAN Tracking

Since WLAN (wireless local area network; standardized wireless network technology based on IEEE 802.11) is available for data-transfer at more and more locations, the progress in tracking devices equipped with WLAN based on the signal strength or run duration establishes new opportunities for user-tracking. This allows reducing the costs because the hardware can be used for both purposes: Data-transfer and positioning.

The first method is based on the fact that signals lose intensity while spreading out in an environment. From measuring the strength at the current position, the distance to the sender can be determined, and triangulation allows for determining the position from several signals. This method assumes a positive correlation between signal strength and distance to the base station. The disadvantage is a high measuring error caused by dynamical changes in the environment, such as shielding of the signal by objects or people.

To reduce this error, the fingerprint method uses signal fingerprints. Each unique fingerprint identifies a unique position in the room. In large rooms, signal patters of the fingerprints recur. Therefore, both methods are usually combined. Obviously, the fingerprint-approach requires the creation of radio-maps in advance, where the signal strengths are measured by hand at different positions. The accuracy of systems like the Ekahau Positioning Engine [Ekahau, 2005] is between 3 and 4.5 meters.

Multi-path distortion caused by radio signals being reflected from walls or people is a big issue in radio-frequency positioning and a source of inaccuracy. Thanks to the short duration of Ultra Wide Band (UWB) pulses, it is possible to filter the reflected signals from the original signal, thus resulting in higher accuracy (about 15 cm). To be tracked, the user must wear a registered UWB tag. The installation of the Ubisense system [Ubisense, 2005] for one room consists of four sensor units mounted at the corners and one control unit. The sensor units' timestamp the signals from all tags present in the room and forward them to the control unit that computes the position of the tags.

Infrared-Based Systems

From its broad dispersion in remote controls, consumer and monitoring electronic, infrared-transceivers (transmitter and receiver together) are cheap, small and energy-saving. On the other hand, IR-technology always need direct intervisibility between transmitter and receiver, the range is limited, and the infrared-signal is reflected by walls, floors and exceedingly by windows. Additionally, interferences with other natural (direct sunlight) and synthetic emitters cut down the bandwidth. The badges send an infrared signal frequently submitting the identifier. The receivers of the signal submit the identifier to a main computer whenever it recognizes an active signal. The accuracy of the detected position depends on the number and dispersion of receivers. The quality of measurements allows tracking single persons in large rooms, checking who is present in the room, or who is sitting next at the table. In large rooms, the system is capable to process several receivers.

Radio Frequency Identification

Radio frequency identification (RFID) is a generic term that is used to describe a system that transmits the identity (in the form of a unique serial number) of an object or person wirelessly, using radio waves. A typical RFID tag consists of a microchip attached to a radio antenna mounted on a substrate. To retrieve the ID and data potentially stored on a tag, a reader has one or more antennas that emit radio waves and receive signals back from the tag. The reader then passes the information in digital form to a computer system.

With the same bandwidth and similar costs compared with infrared, the most important advantage of radio frequency based technology is the needlessness of the direct intervisibility between sender and receiver. Since RF-signals pervade most materials of building, the range is superior (up to more than hundred meters). On the other hand, reflection and reduction of the signal strength affect the accuracy adversely (1 to 3m).

3 Towards Natural Interaction

Despite the fact that multimodal interaction is a relative young research area, it is already quite diverse. This is because of the large range of interaction modes, communication channels and applications that are under investigation. We are focusing on the type of interaction where nomadic users access information, transaction and support systems at the right time and place, and through a range of input and output devices.

Recent multimodal applications (e.g. [Cohen *et al*, 1997], [Eccher *et al*, 2003]) have especially achieved success in support for sensing the multimodal user interaction. This research results in high-end abilities for recognition and synthesis in common modalities such as speech and handwriting. Additionally, current architectures focus on synchronizing events coming from different devices, such as keyboard, mouse, microphone, etc., allowing flexible handling multimodal interactions.

So far, all multimodal services and demonstrators have been designed and built by trial and error. This was inevitable, because of the general lack of understanding of the basics of multimodal interaction. Components that relate to semantic interpretation are far from being standardized. Examples include the fusion of input signals into symbolic representations, representations of the system's interpretation of the overt and covert intentions of the user, dialogue management strategies, and representation of the output information in a way that is contextually appropriate. As a result, services tend to come with idiosyncratic interaction styles, which must be learned from scratch by new users. This confronts prospective customers with long and steep learning curves, and it severely reduces the sustainability of systems and services.

Research must ameliorate this state-of-affairs significantly, by focusing on smart sensor combinations for flexible natural interaction with human users. The goal is to exploit natural human-human communication and interaction patterns for their use in human computer interaction. The interaction between humans and computers happens to be implicit or explicit, whose distinction can be found in the next section.

3.1 Explicit and Implicit Interaction

Apparently *explicit interaction* compasses any kind of interaction with the system, with which the user emphatically expresses desires to the system. The user might communicate such explicit commands in various ways, e.g. with gestures like pointing or voice commands.

If we speak about implicit interaction, we rely on the understanding of the term implicit as used in the paradigms of implicit Learning [Reber, A.S. 1989] [Buchner, A. 1992], as a process occurring inescapably and automatically in the absence of conscious, reflected learning strategies and leading to an abstract representation of knowledge. Implicit process are involved whenever peoples knowledge exceeds what could be concluded from their primary task (plausibility criteria) or from what they could report (operationalisation criteria).

On the other hand, *implicit interaction* summarizes what Nielsen [1993] already called *non-command user interfaces*, and what Oviatt [2000] called *passive modes of interaction*. The idea is simply to use an implicit non-verbal natural behaviour and to extract that information for interaction.

Since methods of implicit interaction tend to be non-obtrusive, they also have the disadvantage of being less reliable than explicit commands. Users tend to intermix unimodal and multimodal interaction; therefore, a combination of several modalities with explicit and implicit interaction can be used for disambiguation. The interpretation and disambiguation of implicit, as well as explicit, interaction must be resolved through context modelling. In contrast to previous work from [Albrecht Schmidt 2000] implicit and explicit parameters are both parts of situational parameters for context modelling.

3.2 Interaction by Approaching Objects

The disadvantage of most approaches especially concerning gesture input is that the action to be performed is often circuitous and hard to learn. But more important gesture interaction is mostly socially obtrusive. Excessive gestures performed without a human counterpart look strange to a casual bystander.

It is a natural behaviour, that if people are interested in something they approach the object to get a closer look to investigate it in more depth. In contrast people, ignore things they are not interested in. Exactly that behaviour (the attitude of curiosity) as explicit as well as implicit interaction with a system can be exploited to learn about the user. The consequence is that the interaction in a physical space can be matched to the informational space for deriving current user preferences and interests. In order to do that the relation between the user and his spatial interaction must be defined.

The aim to use a spatial model as a main source to update and refine the user history requires that the relation between user and spatial model is mapped on a domain model. The domain model describes and classifies objects in the domain and which information they contain, as well as the relation to the user model. The spatial model describes the physical environment of the user and the location of the domain objects in the physical space. The user model describes the knowledge, the interests, and the personal preferences of the user. The users' interactions with the information system and their movements in physical space are automatically evaluated to update the user model.

4 Application Examples

We aim at taking advantage of natural interaction with multimodal systems and go beyond observing input from the different modalities to also integrate the recognition and interpretation of meaningful user-related and environmental parameters. Taking the human computer interaction of a visitor in a museum and a mobile worker in a warehouse as an example we show how we seize the suggestion of Oviatt to develop a blended interface style that combines explicit and implicit interaction methods. Both projects have in common that there exists a relation between the user and her spatial movements, which has to be defined. In more detail, the user's interaction with a physical space corresponds to the user's interests.

Technically speaking, the user's interactions with the information system and her movements in physical space can be evaluated and contribute to automatically refine the user model. The user model describes the knowledge, the interests, and the personal preferences of the user. By overlaying the physical space with an information space and tracking the user's movements in both spaces current user preferences and interests can be derived.

This rather abstract description of how motion can be interpreted by the system as an interaction is now concretized through two specific examples. The following paragraphs describe one completed and one ongoing project of the Fraunhofer Institute FIT making use of the user's movement in the spatial environment as an interaction means.

4.1 The LISTEN Project

In October 2003 the LISTEN system was applied for the visitors of the August Macke art exhibition (see Figure 2) at the Kunstmuseum in Bonn, in the context of the "Macke Labor" [Unnützer, 2001]. Combining high-definition spatial audio rendering technology with advanced user modelling methods creates audio-augmented environments [Zimmermann and Lorenz, 2005]. Visitors are immersed in a dynamic virtual auditory scene that consistently augments the real space they are exploring. The physical environment is augmented through a dynamic sound-scape, which users experience over motion-tracked wireless headphones for 3-dimensional spatial reproduction of the virtual auditory scene.

In order to guarantee an accurate immersion of the user, the high-quality camera-based A.R.Tracking system using retro-reflecting, mostly spherical markers for marking the visitor to be tracked was installed. This tracking system is able to measure the positions of 8 users simultaneously with an absolute accuracy of the head position of about 10 cm. The minimum granularity of the orientation angle comes to 5 degrees.

Figure 2 The LISTEN System applied at the Kunstmuseum in Bonn

While using the LISTEN system, visitors automatically navigate an acoustic information space designed as a complement or extension of the real space [Eckel, 2001]. A sophisticated auditory rendering process takes into account the current visitor's location and head orientation in order to seamlessly integrate the virtual scene with the real one. Speech, music and sound effects are dynamically arranged to form an individualized and situated soundscape offering exhibit-related information as well as creating context-specific atmospheres. Auditory icons providing landmarks in the virtual environment navigation were inserted in the audio presentation to make the user aware of the interaction with the environment. We intended to use stereotypes as an internal model for refining the system and to gather more significant information about the user. Providing stereotypes that are meaningful, easy to detect and to revise without disturbing the visitor was the main challenge.

Several kinds of common behaviour can be identified with people walking through museums (e.g. clockwise

[Oppermann and Specht, 2000]). For mobile users, *travelling*, *wandering*, and *visiting* were seen as three ways to qualify the essence of mobility [Kristoffersen and Ljungberg, 2000]. In comparison, we see travelling as goal-driven relatively fast movement towards a certain destination, wandering as slow sauntering around from one artefact to the other, and finally we split up visiting into two sub-categories depending on whether the visitor is focusing one specific artefact or not. In order to form the basis of adaptive system behaviour, the system accomplishes therefore different audio presentations depending on four motion styles of the user:

- Sauntering around
- Goal-Driven
- Standing, focused
- Standing, unfocused

The interpretation of the user's motion style in combination with the location determinates the presentation style and facilitates the pre-filtering process of relevant sound pieces. If the user stands still focusing on the object, object-dependent information is presented. If the visitor moves slowly not being focused on one specific object, a zone-dependent, more general presentation starts. Finally, the selection of one specific sound piece to be played depends on the user's history of already known sounds and personal interests.

The final evaluation touched upon aspects of interaction as well. The compositional structure of the "Macke Labor" allowed a more distinctive analysis of visitor's acceptance of the introduced forms of interaction. We asked the evaluated persons how they personally experienced the activation of auditory information (technological functionality), emphasizing that they could give several answers: 68% of the 699 evaluated persons acknowledged that the "auditory information seemed to have been activated depending on ones physical movement/position"; only 4% marked "there was no comprehensible connection between ones movement/position and the activation of auditory information." There actually were more than 100 personal remarks on this question. Some people described their experience in terms of having to get used to the system first. We collected statements of the following kinds: "I realized very late how I should behave" or "the longer I moved in the system, the better I could cope with it" as well as positive self-observations such as "one can learn quite easily how to navigate the textual segments" [Zimmermann and Lorenz, 2005].

4.2 The MICA-Project

The project on multi-modal interaction in context-adaptive systems (MICA) is a project on behalf of and in co-operation with SAP. The project started in December 2004 running until November 2006. The goal of MICA is to support workers at their workplace in a natural and unobtrusive way and to proactively help them completing their daily tasks.

In a first phase MICA will be implemented in a warehouse scenario to support warehouse workers in the picking process. The warehouse scenario poses a real challenge for multi-modal interaction. In a warehouse the workers are working with their hands thus requiring hands-free support. The environment is often very noisy and the light conditions might change as well so the interaction is forced to use different modalities according to the needs of the current situation and task. Warehouse men often have to work under time pressure requiring a very responsive system.

The goal regarding interaction in MICA is to provide a combination of explicit and implicit interaction methods in blue collar environments. Unlike the IBM MAGIC system, which uses gaze tracking for the prediction of cursor movement [Paterno, 2002], we favour the combination of speech and pen input with user movement in a physical environment.

Figure 3: Interface of the tablet in MICA

In contrast to the LISTEN project we face situations in which the spatial relations of objects to each other are changing dynamically and need to be monitored and interpreted in real time. In MICA we will combine a low precision WLAN tracking with fine grained UWB Tracking, RFID and camera based motion tracking to be able to determine the spatial and situational parameters.

In the warehouse scenario the system will help the worker to negotiate their way and to work without any problems. They are equipped with a tablet-PC mounted to the trolley, a Bluetooth Headset and tracking tags. A concrete problem where the warehouse-men might struggle in finding the correct location of an item in the warehouse will be identified by analyzing the current task and tracking data and per calculating possible help measures. The system would present an obtrusive hint like a light bulb on the graphical interface and a discreet sound. Only on user's demand, the system would present the help in the correct modality e.g. in a noisy environment on the screen of the tablet, or under bright light conditions via headset. The worker will be able to interact explicitly by approaching objects e.g. if he approaches a package in the incoming goods zone he will receive the information to which shipment the package belongs and if there is a connected task to that package: '20 cartridges to A20'. But also when the worker implicitly passes an object to be picked and approaches another one this information will be processed and he will receive information guiding him back and helping him to find the correct item, or if the approached object is part of another delivery, he will receive information about that object.

As can be seen from the described scenario, the system will be able on the one hand to identify situations in which help might be needed and react on implicit clues in the interaction like stumbling or search behaviour. On the other hand the worker will be able to interact explicitly with the system by approaching objects in the shelf. In particular the combination of implicit and explicit interaction on various modalities will lead to natural blended interaction. It is essential though to improve our understanding of the interactive capabilities that are most important for an automated system to conduct a natural multimodal dialogue.

5 Conclusions and Future Work

In this paper we presented our approach to enhance current multimodal applications to consider more than speech and gestures. From our research in context-aware system development we concluded to integrate other non-intrusive modalities to support natural blended interaction. In particular, the movements of the user observed by a tracking system provide valuable implicit and explicit feedback to the system.

The two described projects show different aspects of the interaction by movement regarding the way user movements are interpreted, and how they are technologically realized. They represent an evolution from implicit to explicit to natural blended interaction combining more and more the spatial interrelations of objects and users.

The substantially improved precision of the fine grained tracking system in LISTEN allowed explicit interaction in the exhibition and the interpretation of the spatial relation of the user to objects in the space. Finally the smart combination of different tracking technologies in MICA shows a first example of this interaction paradigm in a blue collar environment. The interaction of warehouse workers in space will allow to dynamically interpreting their spatial interrelation to each other.

Users positively notice a clear interaction with the system only through their movements. By using the body as an interface the user recognizes the interaction means very fast and is able to intuitively use these mechanisms. The presented projects come closer to the goal of intuitive interaction. The handling should need neither complicated instruction manual nor introduction and should meet all communicational constraints in a social environment: the offered interaction has to take into account the context avoiding the need to navigate through extensive dialogue steps or menus to get to the currently needed information. Appropriate context modelling though is needed to make a major step in intuitive user interaction and proactive information supply adapted to situation, location, task and user.

The MICA project will make an extensive use of user movements. This project is just finishing the requirement acquisition phase with structured user interviews, scenario and mock-up validation and observations on the filed and currently setting up the hardware and starting the specification and implementation phase. A first prototype will be available at the end of this year, which will be evaluated in extensive user tests.

6 ACKNOWLEDGMENTS

LISTEN was funded as an EU-IST project with the contract No.1999-20646
MICA is supported by SAP Research, Contract Nr. DE-2004-044

References

[Brewster, 2002] S.A. Brewster. Overcoming the Lack of Screen Space on Mobile Computers. *Personal and Ubiquitous Computing*, 6(3):188-205, 2002.

[ARTracking, 2005] A.R.T. GmbH – Your Expert for Infrared Optical Tracking. Web Resource: http://www.ar-tracking.de/,.

[Buchner, A. 1992] Buchner, A. Implizites Lernen: Probleme und Perspektiven. München: Psychologie Verlags Union.

[Cohen *et al*, 1997] P. Cohen, M. Johnston, D. McGee, S. Oviatt, J. Pittman, I. Smith, L. Chen, and J. Clow. Quickste: Multimodal interaction for distributed applications. *Proceedings of the Fifth ACM International Multimedia Conference*, New York, NY: ACM Press, pp. 31-40, 1997.

[Eccher *et al*, 2003] C. Eccher, L. Eccer, D. Falavigna, L. Nardelli, M. Orlandi, and A. Sboner. On the usage of automatic voice recognition in a web based medical application. *Proceedings of ICASSP 2003*, Hong Kong, 2003.

[Eckel, 2001] G. Eckel. LISTEN – Augmenting Everyday Environments with Interactive Sound-scapes. *I3 Spring Days Workshop "Moving between the physical and the digital: exploring and developing new forms of mixed reality user experience"*, Porto, Portugal, 2001.

[Ekahau, 2005] Ekahau - The Most Accurate Wi-Fi Positioning. Web Resource: http://www.ekahau.com/, 2005.

[Hightower and Borriello, 2001] J. Hightower and G. Borriello. Location Systems for Ubiquitous Computing, *Computer Issue on Location Aware Computing*, 34(8):57–59, 2001.

[Kristoffersen and Ljungberg, 2000] S. Kristoffersen and F. Ljungberg. Mobility: From stationary to mobile work. In K. Braa, C. Sorensen, and B. Dahlbom (eds.), *Planet Internet*, Studentlitteratur, Lund, Sweden, pp. 137–156, 2000.

[Nielsen, 1993] J. Nielsen. Noncommand User Interfaces. *Communications of the ACM*, 36(4): 83 – 99, 1993.

[Nuamah, 2004] J. Nuamah. Determining the User's Location based on Feature Extraction and Matching of Images. Master Thesis, Fraunhofer FIT, Sankt Augustin, 2004.

[Oppermann and Specht, 2000] R. Oppermann and M. Specht. A Context-sensitive Nomadic Information System as an Exhibition Guide. *Proceedings of the Handheld and Ubiquitous Computing Second International Symposium, Bristol*, UK, pp. 127–142, 2000.

[Oviatt, 1999] S. Oviatt. Ten myths of multimodal interaction. *Communications of the ACM*, 42(11): 74 – 81, 1999.

[Oviatt, 2000] S. Oviatt. Multimodal interface research: A science without borders. *Proceedings of the 6. International Conference on Spoken Language Processing*, 2000.

[Paterno, 2002] F. Paterno, editor. Mobile HCI 2002, LNCS 2411. Springer-Verlag Berlin Heidelberg, 2002.

[Reber, A.S. 1989] Reber, A.S. Implicit learning and tacit knowledge. Journal of Experimental Psychology: General, 118, 219-235.

[Schmidt, A. 2000] Schmidt, A. Implicit Human Computer Interaction Through Context. Personal Technologies Volume 4(2&3), June 2000. pp191-199

[Ubisense, 2005] Ubisense – The Smart Space Company. Web Resource: http://www.ubisense.net/, 2005.

[Unnützer , 2001] P. Unnützer. LISTEN im Kunstmuseum Bonn, *KUNSTFORUM International*, 155:469-470, 2001.

[Zimmermann and Lorenz, 2005] A. Zimmermann and A. Lorenz. Creating Audio-Augmented Environments. *Journal of Pervasive Computing and Communication*, 1(1): 31-42, 2005.