Migrating Characters: Effective User Guidance in Instrumented Environments

Dissertation
zur Erlangung des Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)
der Naturwissenschaftlich-Technischen Fakultät I der Universität des Saarlandes

vorgelegt von

Michael Kruppa

Saarbrücken,
2006

Datum des Kolloquiums:
5. Mai 2006
Dekan:
Prof. Dr. Thorsten Herfet
Vorsitzender:
Prof. Dr. Philipp Slusallek
Gutachter:
1. Prof. Dr. Dr. h.c. mult. Wolfgang Wahlster
2. Prof. Dr. Elisabeth André
Akademischer Beisitzer:
Dr. Jörg Baus

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides Statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Diese Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in anderen Prüfungsverfahren vorgelegt.

Saarbrücken, April 2006

Acknowledgements

I want to thank the DFG (German Research Foundation) for the two and a half year full scholarship that made this research possible, and also the associated “International Post-Graduate College Language Technology and Cognitive Systems”.

Most of all, I would like to thank Professor Wahlster for giving me the opportunity to work and study at his chair as well as at the DFKI (German Research Center for Artificial Intelligence) and especially for his intense supervision during the final stages of this work. Working at his chair was always a joy and I’m very grateful for the many supporting words and helpful hints from my colleagues. I would also like to thank Elisabeth André for being the second evaluator of this work.

Very special thanks go to Jörg Baus, not only for reading through the very first complete draft of this work, but also for showing me everyday that sarcasm and black humor are also known in the Saarland region. I would also like to thank Antonio Krüger for supporting me in both professional and personal matters and for being such a diehard visionary. Furthermore, I would also like to thank Judy Kay and Bob Kummerfeld for the very nice time I could spend at Sydney University and for their advices regarding my research. I’m also very grateful for the wonderful character designs and layouts provided by Peter Rist. Special thanks go to William Niu and Roman Müller for proof reading my thesis.

I would also like to thank my family, which, even though far away, was always supporting me in any possible way. And I do thank Susanne for her incredible patience, her moral support and especially for the kind of optimism she spreads everyday.

I dedicate this work to my little son Sidney who, with a single smile, can always make me forget all my troubles and bring the sunshine in my heart.

Michael Kruppa, April 2006

Wo kämen wir hin,
wenn alle sagten, wo kämen wir hin,
und keiner ginge, um zu sehen,
wohin wir kämen,
wenn wir gingen?

Kurt Marti

Kurzzusammenfassung

Die vorliegende Arbeit beschäftigt sich mit dem konzeptuellen Entwurf und der technischen Realisierung von virtuellen Charakteren, die im Gegensatz zu bisherigen Arbeiten auf diesem Gebiet nicht auf den Einsatz in virtuellen Welten beschränkt sind. Der vorgestellte *Migrating Character* Ansatz erlaubt virtuellen Charakteren vielmehr in der physikalischen Welt zu agieren und zu interagieren. Verschiedene technische Lösungen, welche es einem *Migrating Character* ermöglichen sich in der physikalischen Welt autonom bzw. in Abhängigkeit vom Benutzer zu bewegen, sind ebenso Gegenstand der Arbeit wie eine ausführliche Diskussion der daraus für das Verhalten des virtuellen Charakters resultierenden Implikationen. Während sich traditionelle virtuelle Charaktere in einer wohl definierten virtuellen Umgebung bewegen, muss ein *Migrating Character* flexibel auf sich ändernde Umgebungsbedingungen reagieren. Aus sensorischer Sicht benötigt ein *Migrating Character* also die Fähigkeit eine sich ändernde physikalische Situation zu erkennen. Basierend auf diesen Daten muss weiterhin eine adäquate Anpassung des Verhaltens des *Migrating Characters* geschehen. Neben einer theoretischen Diskussion der notwendigen Erweiterungen eines virtuellen Charakters beim Übergang von virtueller zu realer Umgebung werden auch exemplarische *Migrating Character* Implementierungen vorgestellt.

Short Abstract

The work at hand deals with the conceptual design as well as with the realization of virtual characters, which, unlike previous works in this research area, are not limited to a use in virtual worlds. The presented *Migrating Character* approach on the contrary allows virtual characters to act and interact with the physical world. Different technical solutions allowing a *Migrating Character* to move throughout physical space, either completely autonomously or in conjunction with a user, are introduced and discussed as well as resulting implications for the characters behavior. While traditional virtual characters are acting in a well defined virtual world, *Migrating Characters* need to adapt to changing environmental setups in a very flexible way. A *Migrating Character* must be capable of determining these environmental changes by means of sensors. Furthermore, based on this data, an adequate adaptation of the characters behavior has to be realized. Apart from a theoretical discussion of the necessary enhancements of a virtual character when taking the step from virtual to real worlds, different exemplary *Migrating Character* implementations are introduced in the course of the work.

Zusammenfassung

Im Zentrum der vorliegenden Dissertation steht die Konzeption und Realisierung einer Architektur für virtuelle Charaktere, welche diesen erlaubt neben der virtuellen Welt vor allem auch in der physikalischen Welt zu agieren. Die technischen Rahmenbedingungen werden dabei durch eine variable Kombination von sowohl mobilen als auch stationär installierten Geräten gegeben. Diese so genannten *Instrumented Environments*, also Umgebungen mit integrierter Aktorik und Sensorik, wie zum Beispiel Displays, Soundsysteme oder Kameras bieten eine Reihe von unterschiedlichen Möglichkeiten um (im physikalischen Sinn) mobile virtuelle Charaktere zu realisieren. Des Weiteren werden die von der Umgebung in Form von Servern bereit gestellten Kapazitäten verwendet um rechenintensive Aufgaben von mobilen auf stationäre Geräte zu verlagern. Den Kern dieser Arbeit bildet der *Migrating Character* Ansatz. Er bietet eine Basis zur Entwicklung von virtuellen Charakteren zum Einsatz in physikalischen Umgebungen.

Ziel der *Migrating Character* Technologie ist die Herauslösung der virtuellen Charaktere aus der virtuellen Welt und stattdessen ihre Integration in die physikalische Welt. Es ist ein generelles Ziel des Forschungsgebietes rund um virtuelle Charaktere den Realitätsgrad und die Glaubwürdigkeit dieser Charaktere zu maximieren. Virtuelle Charaktere werden üblicherweise durch zwei Komponenten repräsentiert, zum einen durch die visuelle Darstellung und zum anderen durch die Verhaltenssteuerung der Charaktere. Beide Komponenten haben einen starken Einfluss auf die durch den Benutzer wahrgenommene Glaubwürdigkeit der virtuellen Charaktere. Die hochgradig realistische Darstellung eines virtuellen Charakters in Kombination mit einer höchst unrealistischen Verhaltenssteuerung wird ebenso wenig als glaubwürdig wahrgenommen wie eine völlig abstrakte visuelle Darstellung des Charakters in Kombination mit einer komplexen und korrekten Verhaltenssteuerung. In der vorliegenden Arbeit wird ein weiterer, oft außer Acht gelassener Aspekt in Bezug auf den Realitätsgrad eines virtuellen Charakters diskutiert, nämlich den der häufig eingeschränkten Rolle, die dieser spielt. Selbst der intelligenteste, realistischste virtuelle Charakter wird kaum als solcher wahrgenommen, wenn er zum Beispiel lediglich die Rolle eines Verkäufers auf einer Internetseite spielt. Damit ein virtueller Charakter als realistisches Abbild eines realen Menschen wahrgenommen werden kann, muss er den Eindruck sozialer Kompetenz vermitteln. Dies ist jedoch nur möglich, wenn der virtuelle Charakter auch an realen Situationen, in denen diese Kompetenz wirklich gefragt ist, aktiv teilnehmen kann. Daher ist es notwendig die virtuellen Charaktere in die physikalische Welt zu integrieren, denn nur dort findet die soziale Interaktion zwischen Menschen und zukünftig vielleicht auch zwischen Menschen und virtuellen Charakteren statt.

Im Gegensatz zu virtuellen Welten, deren Ausmaße und interne Zustände vom Entwickler definiert und jederzeit präzise abgefragt werden können, bietet die physikalische Welt einen fast unbegrenzten Freiheitsgrad und eine Fülle an Informationen, die kaum zu erfassen, geschweige denn vollständig zu verarbeiten ist. Das menschliche Gehirn hat im Laufe der Evolution hervorragende Mechanismen entwickelt um vermeintlich irrelevante von relevanter Information zu unterscheiden. Folglich müssen virtuelle Charaktere, welche in der Lage sein sollen sich autonom durch physikalische Umgebungen zu bewegen und mit diesen zu interagieren, über die notwendige Sensorik verfügen, um zum Beispiel Objekte und Hindernisse zu erkennen. Außerdem sollten sie Strategien anwenden können um einem solchen Objekt auszuweichen. Um menschliches Verhalten zu imitieren, sind weitergehende sensorische Daten unerlässlich. Ein virtueller Charakter sollte in der Lage sein, soziale Situationen zu erkennen und sein Verhalten entsprechend adaptieren können. Der *Migrating Character* Ansatz berücksichtigt sowohl die Notwendigkeit sensorischer Daten als auch deren Verarbeitung.

Die technischen Probleme bei der Realisierung des *Migrating Character* Ansatzes wurden in einer Reihe unterschiedlicher Projekte exemplarisch gelöst. Es wurden verschiedene Techniken entwickelt, welche einem virtuellen Charakter die Fortbewegung im physikalischen Raum ermöglichen. Unterschiedliche Sensorik ermöglicht den verschiedenen Charakteren die Wahrnehmung des aktuellen Zustandes der Umgebung. Basierend auf den wahrgenommen Umgebungszuständen und den momentanen Zielen des virtuellen Charakters werden mit Hilfe eines regelbasierten Ansatzes flexible Lösungsstrategien gefunden. Die entwickelten virtuellen Charaktere wurden in unterschiedlichen Szenarien getestet, so zum Beispiel in einer Museumsführer Anwendung und in einer Shopping Guide Umgebung. Jeder exemplarisch realisierte *Migrating Character* ist in der Lage, Benutzer durch physikalische Umgebungen zu begleiten und gegebenenfalls auch die Führung zu übernehmen. Außerdem verfügen die *Migrating Character* über spezielle Techniken und Strategien, die es ihnen ermöglichen, den Aufmerksamkeitsfocus des Benutzers auf bestimmte physikalische Objekte zu lenken. Hierbei wurden menschliche Verhaltensweisen adaptiert. Beispielsweise kann ein *Migrating Character* verbal auf ein Objekt referenzieren. Die Refenzierung kann aber auch in Kombination mit einer Geste erfolgen. Um die Referenz weiter zu verdeutlichen kann der virtuelle Charakter sich auf das Objekt zu bewegen. Im Rahmen einer Benutzerstudie wurde die Eignung der *Migrating Character* Technologie zur Aufmerksamkeitssteuerung der Benutzer positiv verifiziert.

Contents

1 Introduction

Lyra and her *daemon* moved through the darkening hall, taking care to keep to one side, out of sight of the kitchen. The three great tables that ran the length of the hall were laid already, the silver and the glass catching what little light there was, and the long benches were pulled out ready for the guests. Portraits of former Masters hung high up in the gloom along the walls. Lyra reached the dais and looked back at the open kitchen door, and, seeing no one, stepped up beside the high table. The places here were laid with gold, not silver, and the fourteen seats were not oak benches but mahogany chairs with velvet cushions.

Lyra stopped beside the Master's chair and flicked the biggest glass gently with a fingernail. The sound rang clearly through the hall.

"You're not taking this seriously," whispered her *daemon*. "Behave yourself."

Her *daemon's* name was Pantalaimon, and he was currently in the form of a moth, a dark brown one so as not to show up in the darkness of the hall.

"They're making too much noise to hear from the kitchen," Lyra whispered back. "And the Steward doesn't come in till the first bell. Stop fussing."

But she put her palm over the ringing crystal anyway, and Pantalaimon fluttered ahead and through the slightly open door of the Retiring Room at the other end of the dais. After a moment he appeared again.

"There's no one there," he whispered. "But we must be quick."

(a short excerpt from *"The Golden Compass"* by Pullmann, 1996)

If the word “daemon” is replaced with “Migrating Character” in the little excerpt from Pullman’s famous novel on the preceding page, the text illustrates a number of central points of this thesis. In the world created by Pullman, each human being is born with a companion, a so-called daemon. This daemon belongs to a single human and is always in close contact with this person, striving to offer as much support as possible in any given situation. This concept, when transferred to the research topic of this thesis, introduces the idea of an omnipresent Migrating Character assisting its “master” in many different situations.

Even though the vision of such a character may still belong to the realm of (science) fiction, it is worthwhile to observe the character’s behavior in the story in order to identify qualities of the character which need to be realized to come closer to the overall vision. Therefore, a deeper analysis of the story in terms of what it takes for a Migrating Character to behave as the daemon does may be helpful:

When Pantalaimon, now being a Migrating Character instead of a daemon in the original story, first appears in the story, he warns Lyra not to make too much noise. It does so by whispering at her, instead of speaking at a standard volume. Hence, the Migrating Character must be aware of the social setting in which the interaction takes place. He must not only sense that Lyra is making too much noise, but he should also be aware of the fact that this is not suitable in the given situation. Therefore, it is necessary to combine sensory information with a knowledge base, allowing the character to infer the potential risk that Lyra is facing and to react accordingly.

As indicated in the following sentence, Pantalaimon is capable of changing his shape according to a given situation. On one hand, the obvious changes in the character’s shape help Lyra to interpret the role which he is currently playing. On the other hand, taking a different shape allows Pantalaimon, for example, to fly instead of walking. Yet another shape may allow the character to fit into the tiniest space or to appear as a very large and impressive giant. Regardless of the actual shape the character takes, it is always clear to Lyra that he is still her Pantalaimon. Hence, it is very important to ensure that a character’s appearance, even though it may change to a certain degree, always keeps specific features which identify it as a unique character.

Finally, Pantalaimon “flutters” ahead alone, leaving through an open door. When he comes back a moment later, he reports to Lyra what he has seen in the other room. The character is hence not only capable of following Lyra through the physical world, he may also move autonomously within a specific range in order to gather information or to point Lyra towards a specific object. While staying together with Lyra, the character’s movements are somewhat passive, since they are directly linked to Lyra’s movements. To allow the character to move autonomously, it is necessary to support active locomotion as well. However, when actively moving away from Lyra, the character needs to ensure that she notices where he went so that she can follow him if she likes, or wait until he comes back to her. Hence, Pantalaimon chooses a form which allows him to quickly move from one place to another but also allows Lyra to follow his movements.

1.1 Aims and Methods

The analysis of the story presented in the previous section already indicates the various problems which need to be addressed in order to allow a Migrating Character to become an omnipresent companion to its user, much like the daemon in *The Golden Compass*. Even though a complete realization of such a character is still a distant goal, one may already see that even the first steps towards this vision demand an interdisciplinary approach combining different research areas (as indicated in Figure 1.1).

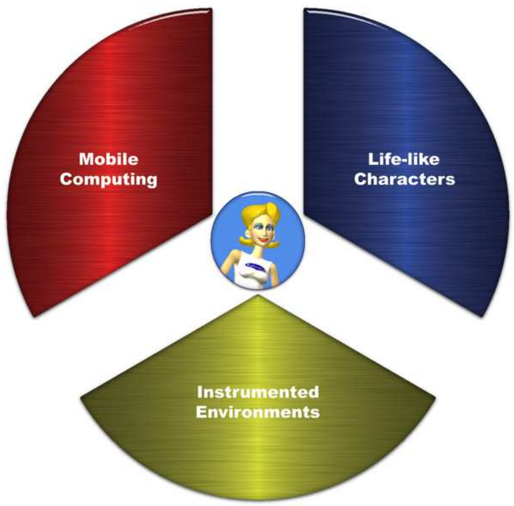


Figure 1.1: Research areas combined by the Migrating Character Technology

Most obviously, we will have to review and adopt methods and technologies from the field of life-like characters. Research in this area typically strives to improve the degree of realism (i.e. believability) of life-like characters by realizing sophisticated behavior models and engines as well as realistic character visualizations, allowing a character to combine gestures, facial expressions and spoken language in order to communicate with a user. In Section 3.7 we will argue that a major limitation which hinders all life-like character implementations so far from reaching a new level of believability, is the limited role and flexibility they play. It is a major aspect of this thesis to explore new methods which will allow a single life-like character to support users in a very flexible way in many different settings, utilizing a wide range of different technologies. Generally speaking:

> *A Migrating Character is a life-like character which is able to migrate from one environment (physical location or device) to another. The migration may be either initiated actively by the Migrating Character itself or the movement may be triggered by a user. In either case, migrating a character means to transfer the characters state and execution from one environment to another. A Migrating Character incorporates knowledge about the physical world around it and it utilizes sensor information to adapt to a changing environment, allowing the character to assist and furthermore to* ***guide and follow*** *a user while exploring both* ***virtual and physical environments****.*

Consequently, we will have to incorporate the field of mobile computing into our research and development plan for the Migrating Characters. Including technologies and methods from this research area will not only allow a Migrating Character to be taken along by its user while exploring physical spaces, but it will also allow the character to sense at least a subset of the physical setting around it (as discussed in Section 2.3).

However, if we want to go one step further in the improvement of a life-like character's flexibility, it will be necessary to include means for autonomous locomotion in these characters as well. As a result, the character will not only be more flexible regarding its own movements, but the autonomous locomotion may also further improve the believability of the character. For example, when referring to physical objects, an autonomously moving life-like character may mimic human behavior to a high degree by moving closer to the particular object, and pointing at the object while referring to it verbally (see Sections 4.2.1 and 6.3 for further details). From a technical point of view, the inclusion of further hardware appliances in addition to a mobile computing system will be necessary in order to realize this technology. The research area of instrumented environments deals with the seamless integration of networked appliances (such as sensors as well as presentation devices, like video and/or audio systems) in the physical environment and thereby offers promising solutions which allow a life-like character to move with a maximum degree of flexibility through physical spaces by utilizing hardware integrated in the environment. From these considerations, we derive several questions and issues that we have to address in order to conceptualize and realize the Migrating Character technology:

- **What are the desired capabilities of a Migrating Character?**

  The example presented in the previous section hints at the major aspects we strive to achieve by realizing the Migrating Character technology. Throughout this thesis we will introduce and discuss the different aspects and features which should be realized in the Migrating Characters so as to allow them to act as competent guides in instrumented environments.

- **What are the preconditions for realizing Migrating Characters?**

  In order to realize the different Migrating Character abilities, we first need to identify certain preconditions. Such preconditions may be, say, the sensory information necessary for the character to perform a specific task, or the specific knowledge required to derive inferences from sensory data. We will introduce and discuss all necessary preconditions for the development of the Migrating Character prior to a discussion of the concrete realization of the Migrating Character technology.

- **What technologies are necessary for specific Migrating Character aspects?**

  A Migrating Character, unlike a traditional life-like character should be capable of moving and interacting in physical spaces. While a life-like character is constituted by a piece of software, the realization of the Migrating Characters also involves the choice and integration of adequate hardware during the development process. We will identify different hardware categories, such as sensors, mobile devices and stationary presentation systems, each of which will be necessary to realize certain features of the Migrating Characters.

- **How can we develop a device-independent character engine?**

  In order to maximize the flexibility of the Migrating Characters, they should not only be capable of working on different classes of devices, but they should also be developed in a so-called device-independent data type. In this way, the effort for the integration of new appliances into a specific scenario may be minimized.

- **How can we maximize a characters capabilities on a mobile device?**

  Mobile devices, as opposed to stationary ones, have to deal with a number of limitations. For example, they have to deal with low-power components resulting in inferior computational capabilities when compared to stationary devices. Equally restricting is the generally limited screen space on mobile device due to the portable nature of such devices. The Migrating Characters, which will have to work on both mobile and stationary devices, have to deal with these limitations in a way that allows maximizing the usability of the character while on a mobile device. We will discuss methods to deal with low computational power by relaying computational tasks to stationary devices and we will introduce design guidelines for Migrating Characters on mobile devices.

- **How can we allow for a consistent character behavior and appearance?**

  As indicated above, the Migrating Characters should work on different appliances belonging to different classes of devices and they should be developed in a device-independent fashion. Following such an approach bears the risk of producing different results on different devices. While this may be acceptable in traditional systems, the use of life-like characters demands a consistent appearance and behavior of the character regardless of the different technical setups. This consistency will allow a user to identify and follow a single character on different devices. We will hence discuss how character consistency may be guaranteed, regardless of the combination of devices in use. This will involve different character layouts for different device classes as well as strategies to minimize the demand for computational power on mobile devices by including additional stationary servers.

1.2 Outline

Apart from the scientific goals discussed in this introductory chapter, we pursue another important goal in writing this thesis: by providing a clear structure, we want to facilitate various approaches for browsing through this work. Each of the following chapters shares the same basic layout, starting with a short introduction presenting an overview on that chapter and its underlying structure, represented by different sections and subsections. At the end of each chapter we present a summary which reviews and subsumes the most relevant points discussed throughout that chapter.

The thesis is subdivided into three major parts, as indicated in Figure 1.2.

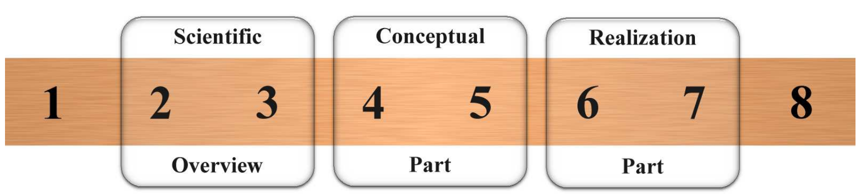


Figure 1.2: Structural overview of this thesis

The first part, starting with the following chapter, provides the scientific background necessary for the development and the understanding of the concepts and technologies described in this thesis. It starts with an overview on basic concepts relevant for our work in chapter 2. Topics discussed in this chapter cover important concepts from each of the different research areas which were combined to allow for the development of the Migrating Characters. The conclusions of this chapter include a discussion of the different aspects and concepts for each of the research areas and how they may complement each other in a beneficial way. This overview on basic concepts is then followed by an extensive discussion of previous and related work in chapter 3. Similar to chapter 2, the inherent structure of this chapter is defined by the different research areas which form the background of our research. The overview is concluded by a summary of typical problems for each of the mentioned research areas. Finally, a discussion of the different approaches towards those problems, based on the reviewed literature, and the introduction of new solutions provided by the Migrating Character technology conclude the first part of the thesis.

The second major part is dedicated to the conceptualization of the Migrating Character technology. In chapter 4 the Migrating Character concept is introduced. The Migrating Characters are entering the realm of mobile computing and the field of instrumented environments. A discussion of novel possibilities when taking the step from life-like characters to Migrating Characters hence follows the introduction of the concept. In addition, prerequisites which need to be met in order to allow the Migrating Characters to realize specific behaviors or features are also discussed. After a summarization, a central aspect of the Migrating Character technology is discussed in chapter 5, namely the coherent design of the Migrating Characters for different device classes and platforms. The main goal of the chapter is to provide general guidelines for the design of Migrating Characters, which should help developers to avoid typical pitfalls when realizing their own Migrating Characters.

Chapters 6 and 7 form the third major part of the thesis and provide an overview on a number of different realized prototypes as well as an extensive report on a user study performed within the context of this research. The different prototypes introduced in chapter 6 illustrate the general feasibility of the Migrating Character technology. Furthermore, they also give an idea of what may be achieved by using this technology and how flexible the Migrating Characters may be in different application scenarios. The user study presented in chapter 7 focuses on one particular aspect of the Migrating Character technology but gives an idea of how other aspects may also be evaluated in a similar fashion.

To conclude the work at hand, a summary of the achieved results and scientific contributions is presented in chapter 8. A consideration of possible future work wraps up the thesis, and represents the punch line of the work at hand.

Scientific Overview on Concepts and Projects Relevant for the Development of the Migrating Characters

2 Basic Concepts

Before we review related work and before we present the main results of this thesis, we need to introduce some basic concepts that we will use throughout this work. We organized this chapter into three sections, focusing on different aspects relevant for the development of the Migrating Characters.

One of the major advantages of the Migrating Characters is their ability to efficiently refer to relevant physical objects. The basic concepts behind object referencing in general are introduced in Section 2.2. Another important feature of the Migrating Characters is their flexibility, especially with respect to relocation in physical space. However, when leaving the standard desktop metaphor, we enter the area of mobile and ubiquitous computing that has its own restrictions and concepts which we will discuss in Section 2.3. Then a summary of specifications and concepts from the area of life-like characters will be presented in Section 2.4. We conclude the chapter with a discussion on the interrelation of concepts between object referencing, mobile and ubiquitous computing and life-like characters as well as problems arising during the integration of them (see Section 2.5).

We will start this chapter by defining the term “instrumented environment” as we will use it frequently throughout this work.

2.1 Instrumented Environments - Term Definition

Quite a number of projects have used the term “instrumented environment” (IE) to describe the framework in which their work has been developed. However, until now there is no clear definition of what the term actually refers to. In this section we will give a short overview on how the term is used in a number of different publications. We will then continue by identifying a common understanding of what IEs are, based on how the term has been used in those projects. Finally we will state a clear definition of how we will use and understand the term throughout this work.

Terrenghi (2005) states that IEs are a subclass within the research area of ubiquitous computing (see section 2.3), defining spaces with embedded technology such that objects of everyday live may be used to display and sense information. The author further explains that within IEs, users may interact in a continuous display of information by handling physical objects which are linked to virtual information and by moving through physical space. According to the author, the fact that the information “migrates into the walls”, where different devices are interconnected in an invisible fashion to the user, constitutes the biggest problem in IEs since it bears the risk for the users to lose both control and awareness of the interaction (this is also often referred to as the “invisibility problem”, see also Carmichael, Kay, & Kummerfeld, 2005).

In (Kray, Krüger, & Endres, 2003), the authors use the example of a modern airport to illustrate their understanding of an IE. The airport scenario described includes very large public boards to display arrivals and departures but also smaller plasma screens at gates which display information on actual flights. In addition, small touch screens are installed throughout the airport building allowing users to access information on airport facilities. Private devices, like Personal Digital Assistants (PDA) or laptops may also be integrated in the scenario by means of wireless LAN access points. The authors speculate that in the near future the amount of displays installed in an IE will constantly increase. Apart from these visual displays, other devices like audio systems and a variety of sensors may be integrated in IEs. The main problem addressed by the authors is that of using all this technology in a combined fashion to improve the overall presentation quality, focusing especially on visual displays.

Another interpretation of IEs is stated by Stahl et al., 2004. According to the authors, such environments incorporate distributed computing power, presentation media as well as sensors. Furthermore, the authors integrate into their understanding of IEs the capability of systems running in the IE to observe and recognize implicit user interactions. One of the main issues, as claimed by the authors, is that of using the data coming from sensors to recognize user interactions in order to infer about a user’s plans and intentions. Based on this inferred data, a user could be proactively assisted while interacting with the IE.

After reviewing these different understandings of IEs, we may summarize that all authors agree on the fact that such environments are packed with technology which often may be only partly visible or even completely invisible. The technology is integrated within the environment, making it less intrusive but also harder to be recognized by the users. It is hence necessary to support users while interacting with IEs in order to maximize the benefit from the provided technology. As stated by all authors, the interconnection of the different appliances within an IE is what really distinguishes it from just an accumulation of fancy technology. Communication among devices and services, and especially also the integration of personal devices into the communication channels, is one of the central aspects of IEs.

Within the scope of this work we define instrumented environments as follows:

> *Instrumented environments augment common areas of everyday life, such as homes or offices, with networked appliances. Computational evaluation mediates between two different types of embedded appliances, namely sensors and actuators, and allows an instrumented environment to sense and interpret information and to produce corresponding reactions.*

2.2 Object Referencing Related Concepts

When humans refer to physical objects or entities, they utilize knowledge about the arrangement of objects in physical space. Even though most people have an intuitive understanding of what the term *space* refers to within this context, there is no exact definition of what is meant when talking about space. However, since the Migrating Characters described in this work try to imitate the way in which humans communicate about spatial arrangements of physical objects, we need to define the terms used throughout this thesis and to give an overview on different approaches related to space.

2.2.1 Space

Most people share a naive interpretation of the term *space*, which defines space as being the physical environment in which we live. According to Freksa and Habel (1990), the term *space* generally depicts structures formed by combining an arbitrary number of elements. The term is used in different ways, depending on whether it appears in a natural sciences- , a humanities- or an everyday context. Freksa and Habel distinguish between *psychological space*, *metaphorical space*, *physical space* and *geometrical space*.

The authors define the psychological space as a concept to describe the way in which biological systems in general perceive space. The way biological systems sense the surrounding environment is extremely variable because it is realized via manifold channels of perception. A human being, for example, is capable of perceiving attributes of the environment visually, acoustically, haptically and olfactorily.

In order to communicate the meaning of non space related concepts, the metaphorical space utilizes know-how about the physical space and attributes of spatial concepts.

The physical space is defined by three positive, orthogonal axes forming a three dimensional coordinate system based on real numbers (this concept is inspired by Newton's definition of the homogenous, three dimensional space). Adding a fourth dimension allows for defining forces of different ranges and intensity. For example, the axioms of movement determine a link between time and physical distance.

Finally, the *geometrical space* is a structure (based on a point set) which is defined by a number of axioms which need to hold for each point within the point set. The basic concepts of point, line, and plane define terms like distance, area, volume and angle.

2.2.2 Spatial Knowledge

When humans acquire spatial knowledge, the information gathered can be separated into three different categories of spatial knowledge: *landmark knowledge, route knowledge and survey knowledge* (see also Lynch, 1960, Tversky & Lee, 1999 and Golledge, 1999).

Landmark knowledge is a representation of visual details of a specific location. Usually, especially striking objects like, for example, huge buildings are subconsciously chosen as landmarks during an exploration phase. However it is also possible to gather landmark knowledge by studying pictures and visual representations of a specific area. Landmark knowledge helps to recognize certain spatial areas by remembering landmarks. The term landmark is used in different ways in literature. While Sadalla, Burroughs, and Staplin (2001) use the term to define reference points which are better memorized than other points, Lynch (1960) understands landmarks as objects which, due to their uniqueness, are chosen as the most distinctive features of different areas. Landmarks are used to structure spatial knowledge. Since these landmarks stand out from their surroundings they are easy to remember and to recognize. Landmarks are useful in many different spatial processes, like for example way finding (Raubal & Worboys, 1999; Lynch, 1960) or object localization (Gapp, 1997). It is actually possible to link other knowledge to landmark knowledge by, for example, giving way instructions which associate certain landmarks with decision points, hence simplifying the task of deciding which path to choose.

While exploring a specific area, humans accumulate spatial experiences. Apart from landmarks, also paths or routes are recognized during this process. This accumulated knowledge is referred to as route knowledge or as procedural navigational knowledge. Knowledge of this type is represented as a sequence of actions and decisions while following a path in a specific environment. Alternatively, it is also possible to acquire route knowledge indirectly by studying a map or listening to route instructions. The main characteristic of route knowledge is the fact that it is gained in a successive manner (as discussed in May, 1992).

The third category of spatial knowledge, survey knowledge, can be either derived from a combination of landmark knowledge and route knowledge or by studying a map. Survey knowledge encodes information about spatial constellations and the specific topology of an area. Survey knowledge has been acquired when a person is capable of building a network of spatial relations between landmarks. This network allows a person to infer relative information about objects, without having physically experienced the path between the objects. A person possessing survey knowledge of an area is capable of finding shortcuts and effectively navigating between different locations. In addition, survey knowledge allows a person to estimate the direction to a landmark (see Montello, 1998).

2.2.3 Spatial Relations

Apart from acting within space, human beings also communicate about space and spatial concepts. By combining verbal utterances and according gestures, humans are capable of referring to specific objects in space. This methodology also allows for complex spatial descriptions. In addition, humans who are capable of writing are also capable of putting down spatial descriptions in written words. A language in general is a means to allow a speaker/writer to construct a mental representation of the described scene in the minds of the receiver (see Fauconnier, 1997). Results of several different experiments (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Kintsch, 1998; Dijk & Kintsch, 1983) indicate that the process of textual comprehension equates with the construction of a mental representation of the described scene. The mental representation is built up while the receiver reads or hears the descriptions. While processing verbal spatial descriptions, which usually consist of propositional statements about the scene to be described, there are two possible ways for a receiver to work with the information. The receiver can either try to remember the accumulated information or he/she can try to visualize images based on the spatial descriptions. Regardless of which technology comes to use, both allow for further processing later on, for example, to infer spatial relations between objects which were not explicitly described.

A very common way to realize spatial references is the use of spatial relations (see Herrmann & Grabowski, 1994). Spatial relations are constructs consisting of three main parts. First of all is the relation itself. The second part is the reference object, which is also-called anchor object. Depending on the type of relation it may be necessary to have more than one anchor object, for example, when referring to an object which is located between two other objects (see also Habel, Herweg, & Rehkämper, 1989). The anchor point defines the origin (ground) for the relation. The "object to be localized" is called the target object and constitutes the third part of a spatial reference. The target object is the object which is spatially related or localized to the anchor object. Schirra (1994) has argued that the whole expression may also be graded with a *degree of applicability* (DA) which represents the degree to which a relation applies to the situation it describes.

2.2.4 Spatial Deixis

Before discussing the meaning of the term *spatial deixis* we will examine the two words which were combined to form the new term.

Deixis is often used in its linguistic meaning, referring to the use of categories and items of lexicon and grammar which are controlled by certain details of the situational context in which the utterances are produced (see Fillmore, 1982; Lyons, 1977). These details, apart from the identity of the participants of the communicative act, also contain their orientations and locations in space, the time at which the utterance is produced as well as any on-going indexing acts in which the participants may be involved in. When speaking of deixis in natural language, research in the linguistic area deals with two different questions:

- How do speakers succeed in anchoring referential acts in space and time by taking into account their current situation?
- What lexical and grammatical means are provided by a given language, dedicated especially to such purposes?

The term *spatial* is a derivation of the term *space* which was already discussed in Section 2.2.1. However, spatial notions in natural language semantics are commonly based on a three dimensional system of coordinates with axes labeled as *up/down*, *left/right* and *front/back* as opposed to the geometrical notion of space with arbitrary zero points and vectors. The *up/down* axis is usually determined by the direction of the gravity force. It is well known that the interpretation of spatial expressions depends on the selected frame of reference (for a more complete discussion of the intrinsic versus extrinsic use of spatial prepositions see (Retz-Schmidt, 1988)). One way to establish a reference frame is to use an intrinsic orientation. In this case the *front/back* axis is determined by the speakers ability to move with a fixed orientation (and is therefore anthropocentric) while the *left/right* axis is essentially egocentric.

In combination, the term *spatial deixis* refers to a subcategory of deixis, namely those aspects involving references to the communication participants' locations in space. In other words, spatial deixis takes the bodies of the participants of the communicative act as reference objects for spatial specification (as discussed in Fillmore, 1982).

2.3 Mobile and Ubiquitous Computing Related Concepts

The term *mobile computing* first appeared in the early 90's when advances in wireless networking technology lead to this new paradigm. Unfortunately, there has never been a generally accepted definition of what exactly the term refers to. Since the field of mobile computing is so diverse, a definition like the one of Forman and Zahorjan (1999) seems to capture the spirit of this research area most appropriately:

> *Mobile computing refers to a paradigm of computing in which users have access to a shared infrastructure independent of their physical location by means of portable devices.*

This definition also indicates how different mobile computing is from a standard desktop computing metaphor. In a typical desktop computing environment for example, even though there exist quite a number of different approaches, the general interaction between user and computer system is realized via a combination of a keyboard and a pointing device for input purposes while standard displays and audio devices realize the opposite communication channel. Whereas in mobile computing, the variety of devices and input/output modalities is immense, ranging from mobile phones with reduced keyboards and small simple displays to portable Tablet PC with larger, high resolution screens and touch pads.

Apart from the different setup of mobile devices, an ever growing number of different wireless communication technologies, as well as an equally fast growing number of services available for mobile users, not only offer new possibilities but also pose problems which were almost unknown when computers were stationary devices with fixed cables. While communication channels on such stationary, cable connected devices are usually very stable, mobile applications have to deal with the fact that wireless communication technology is often less reliable, resulting in decreasing bandwidth and, in the worst case, in a complete loss of the communication channel. Any mobile application trying to convey the impression of a professional, stable system hence needs to provide solutions for situations in which necessary communication channels are not available.

On the other hand, mobile computing offers a wide range of new application possibilities. The fact that mobile computing refers to scenarios in which users and devices constantly move from one location to another, does not only result in technical problems, but also in completely new classes of services. With mobile computing it is possible to integrate computer services into the physical world. If a mobile computing application is capable of sensing the location of the user (or usually the location of the device the application is running on), it is possible to offer location-based services. In this way it is, for example, possible to logically combine information access on a mobile device with the spatial layout of the environment and the arrangement of physical objects. Location awareness is hence one important feature of mobile applications (different technologies to determine positional information in mobile computing applications are discussed in section 2.3.2). Positional information, however, is useless, unless a mobile computing application can draw conclusions from this information. Hence, depending on the type of application, spatial knowledge is necessary for a mobile computing application to determine the actual context of the user (a discussion on the term *context* follows in Section 2.3.1).

The same problems and possibilities are also faced by researchers in the area of ubiquitous computing. In (Weiser, 1991), the author describes his vision of the computer for the 21st century, a paradigm which he refers to as ubiquitous computing. The idea behind this new computing paradigm is the one of a world in which desktop computers are replaced by computing embedded in physical objects without interfering or hindering the current functionality of such objects. Instead of interacting with visible, desktop-based computers, ubiquitous computing devices would be small enough to be actually invisible inside physical objects. These integrated computing devices would than enhance the original functionality of the physical device they inhabit. In the author's vision, people would still do their work assisted by computer technology. However, as in contrast to today's computing metaphors, users would no longer have to focus on the computing devices.

2.3.1 Context

Most people have an intuitive understanding of what is meant by the term *Context*, even though it is very hard to actually describe or define the term. In general, everything that happens in the world, no matter whether it is noticed by humans or not, takes place within a certain context. From this fact, a general description of the term context is derived, which defines context as a number of facts and circumstances which describe a situation, an event or an environment. In the research area of computer sciences, context is defined in many different ways. Brown (1996) defines context as a number of elements in the vicinity of the user which are aware to the computer while Ward, Jones, and Hopper (1997) understand context as the state of the environment in which an application runs. In contrast, Rodden, Cheverst, Davies, and Dix (1998) allude context to the setup of the software and for Franklin and Flaschbart (1998), context is solely related to the user and not the computer system. Another, different view is provided by Hull, Neaves, and Bedford-Roberts (1997) in whose definition, the whole environment is included, but only certain aspects of the current situation are understood as context.

Even though the definitions discussed above are very diverse, they have one thing in common: Humans may easily understand and apply these definitions however they are not easily processed by computer systems. While developing context sensitive services, different research groups have suggested definitions for the term context which better fit in the computer science area. The definition of Schilit, Adams, and Want (1994) and respectively that of Schilit and Theimer (1994) is located in the area of mobile and ubiquitous computing and distinguishes between three different types of context, all relevant in the specific application area:

- **Computer context:** The computer context subsumes information about available technical resources in the environment (for example, printers and workstations) as well as information regarding available network resources and corresponding information about bandwidth or cost of communication.
- **User context:** The user context holds information about a user of a system, her interests and location within the environment as well as information about all other surrounding people.
- **Physical context:** The authors define the physical context as a collection of information about properties of the physical environment, including information about lighting conditions, ambient temperature or traffic conditions.

Pascoe (1998) understands context as a subjective concept which is defined by the entity perceiving the environment. This definition is highly dependent on the particular interests of the entity. In one case, the location may be important, while in another case, the temporal perspective may be relevant. Even the emotional state of another person may be considered relevant and hence become part of the context. In general, Pascoe defines context as a subset of physical and conceptual states which are of interest to a specific entity in a particular situation.

Chen and Kotz (2000) add a fourth dimension to the definition of context, the so-called time context. The time context includes information about the time of day, date and time of year. This data may influence the behavior of a computer system, it may for example modify the visual appearance on a screen, depending on the day of time and corresponding lighting conditions. In addition, the authors also mention the collection of context data over a period of time.

A design practice oriented definition of context, located in the area of mobile applications, is suggested by Raptis, Tselios, and Avouris (2005). Similar to the approach of Chen and Kotz, the authors subdivide context into four different categories:

- **System context:** The system context reflects the structure of all interconnected devices and corresponding services which form the basis of the system. The system context is especially related to the way in which the systems functionality is distributed across the different devices and the devices' awareness of each other.
- **Infrastructure context:** The infrastructure context refers to information regarding the way which the different devices and their integrated applications communicate with each other across different communication channels. The infrastructure context sheds light on the quality of information, namely how timely and accurate the information of a node in the system is.
- **Domain context:** The domain context is related to the adaptability of a system to specific characteristics of particular users. The domain context influences the way the system alternates the quantity and type of information that is presented to a user, based on the needs of each individual user.

- **Physical context:** The physical context is subdivided into the categories location, mobility and population. While location describes the ability of a mobile device to locate itself in space, mobility describes the devices agility (for example fixed, mobile or autonomous), the devices relation to other devices (free or embedded) and the number of people which may concurrently interact with the device. Finally, the population category describes the sort of bodies that populate the application space.

The presented definitions for the term context in this section show a wide range of variety. Regardless of which definition is chosen, mobile applications are highly dependent on the physical context. In order for these mobile applications to be capable of taking appropriate actions due to a changing physical context, these systems need to be aware of their own location as well as the location of relevant objects in the environment. In the following sections we will give an overview on different technologies allowing for position detection by mobile devices and we will discuss appropriate ways of representing spatial knowledge in computer systems.

2.3.2 Positional Information

Positional information does not only refer to a person's or an object's relative or absolute locations. In addition, the positional information also includes such factors as viewing direction (if applicable) and body orientation. Information about speed and acceleration is also part of the positional information. Up to present there is no single sensor available providing all the information at once and in all cases and situations. Hence, especially in general purpose applications, it is necessary to combine the data of several different sensors in order to maximize the quality of the positional information available for the application.

In order to provide location-based services in mobile and ubiquitous computing applications, determining positional information of both users and devices is a necessary first step. There are several different methods to accomplish this goal. We will introduce the most common position detection techniques and we will discuss the benefits and disadvantages of each technology in the following sections. In general, the different technologies can be categorized in either indoor or outdoor position acquisition techniques.

2.3.2.1 Outdoor

The most common technique to detect the current location in outdoor scenarios is the *global positioning system* (GPS). The GPS technology is satellite-based and is formed of at least 24 satellites orbiting at about 12,000 miles above the surface. The low orbits of the different satellites are organized with the goal to ensure that at any given time and location, a surface-based GPS receiver should be within transmission range of at least four satellites. This is necessary, since the calculation of the receivers location is based on runtime difference of the satellite signals. Each GPS receiver features an almanac with information on each satellites position at any given time. The satellite signal includes a unique identification code and a very precise time stamp (each satellite features an atomic clock). Based on the pre-calculated positions of each satellite and the time difference between the different satellite signals, the receiver is capable of calculating its own location via triangulation (this calculation is illustrated in Figure 2.1).

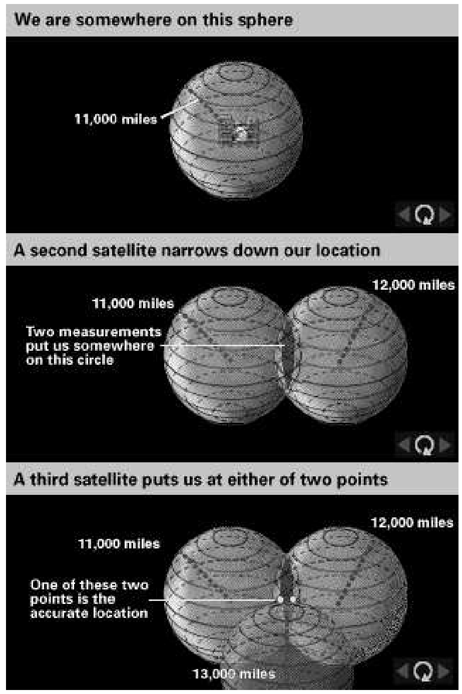


Figure 2.1: GPS triangulation using one, two or three satellites (source: Trimble Navigation Limited, visited 2005)

However, since the internal clock of the GPS receiver cannot be as precise as the ones of the satellites, the GPS receiver needs the signal of a fourth satellite in order to correct its internal timing error. If the internal time of the GPS receiver were correct, the sphere corresponding to the fourth satellite would intersect with the former three spheres in a single point. If the timing is not correct, there will be an offset between the first three and the fourth sphere. This offset can be used to recalibrate the internal clock of the GPS receiver. The accuracy of GPS depends on the number of satellites whose signals are received simultaneously. In order for the GPS receiver to receive the signal of the satellite, the receiver needs to “see” the satellite, i.e. there has to be direct line of sight between transmitter and receiver. However in reality, the GPS signals are often obstructed by buildings, dense vegetation or even because of clouds during bad weather. These limitations also imply that, the GPS system does not work inside buildings, since the signal is obstructed by the walls of the building.

Also based on triangulation is another location technology which is using the network cells of cellular phone companies. The benefit of this approach is the fact, that these network cells also reach most inside areas of buildings. Since the locations of the senders are fixed, the position calculation is very easy. The main problem of this approach however is the fact, that these network cells sometimes have large ranges (between 500 meters and several kilometers, depending on the topology of the surrounding area), resulting in a very low positional precision. An alternative to cellular phone network cells is to use network cells of short range wireless communication protocols like *Wave LAN* and *Bluetooth*. Due to the far smaller cell sizes, the positional data acquired in this way is more precise. However, as compared to cellular phone network cells, the coverage of these cells as of today is insufficient. In order to work properly, a fully developed infrastructure of senders would be necessary.

The aforementioned position detection technologies all share a common aspect. In order to detect its own location, a positioning device depends on other, external technology. If the communication with this external technology fails, the whole position detection system fails.

A very different approach is based on *dead reckoning*. Instead of determining absolute positions, the dead reckoning technology is based on the idea of starting at a fix point (whose position is known to the position detection system) and accumulating relative movements in order to estimate the current position. In order to do so, it is necessary to know both heading and speed at any given time. This can be realized by using an electronic compass to determine the movement direction, and accelerometers to determine the acceleration and the corresponding movement speed. The main problem of this approach is based on the fact that any measurement error will result in a constant positional error, which can neither be detected nor corrected.

2.3.2.2 Indoor

As mentioned in Section 2.3.2.1, the GPS does not work properly inside buildings, since the communication between senders (i.e. satellites) and receivers is hindered by the structure of the building. However, it is still possible to use GPS inside buildings by setting up additional GPS senders around the building. The indoor GPS senders determine their own position and internal time by using the outdoor, satellite-based GPS. Based on these data, several senders at different locations emit a signal very similar to the satellite signals. An indoor GPS receiver can receive these signals inside the building and calculate its own location based on the same algorithm as of the original GPS. The benefit of the indoor GPS lies in the fact that, once the system is set up, it can cover a whole building. However, monetary prices for indoor GPS senders and installation are very high.

Wave LAN and Bluetooth position detection systems also work inside buildings. However, due to the radio frequency used for these communication techniques, the signal may be influenced by walls and windows. Furthermore, as experiments have shown, the signal strength may also be influenced by bodies of humans or animals – due to the high amount of water within these bodies – within the range of the sender. A benefit of both Bluetooth- and Wave LAN-based indoor navigation systems is the inherent back-channel provided by the system. This is of special importance in scenarios, where the determined positional information is not only used locally, but is also processed by external services (e.g. a scenario with autonomously moving devices, controlled by a central system).

Dead reckoning, combining an electronic compass and accelerometers, is a possible solution for indoor tracking. However, the expected error rate is higher, due to the fact that the electronic compass is very sensitive to metallic objects. Furthermore, electromagnetic fields (which are very common nowadays inside most buildings due to the use of electronic devices like computers, television etc.) cause strong interference and distortion within the electronic compass.

Another electromagnetic positioning technology is based on small tags emitting radio signals. These so-called RFID (Radio Frequency IDentificaiton) tags are available in two versions. While

Technique	Scenario	Type of Data	Limitations
GPS	Outdoor	Position	Low Precision in urban regions
Cellular phone networks	Outdoor/Indoor	Position	Very low precision in rural regions
Indoor GPS	Indoor	Position	Very high installation costs and effort
WirelessLAN	Outdoor/Indoor	Position	Coverage, sensitive to presence of human bodies
Bluetooth	Outdoor/Indoor	Position	Coverage, sensitive to presence of human bodies
Infrared beacons	Indoor/Outdoor	Position/Heading	Needs a direct line of sight in order to work
RFID tags	Indoor/Outdoor	Position	Coverage, sensitive to presence of human bodies
Electronic compass	Indoor/Outdoor	Heading	Sensitive to metallic structures
Electronic accelerometer	Indoor/Outdoor	Acceleration	Low precision, data only available while moving

Table 2.1: Comparison of different positioning techniques

passive RFID tags emit a signal (usually a unique ID) when entering the range of specialized antennas, active RFID tags have an internal power source and emit a constant signal. The range of the passive RFID tags is small, which is why they are mainly used to detect their presence in a narrow range, for example in anti-theft systems in shopping malls. Active RFID tags have a bigger range and may be detected by small antennas which are also available for several mobile devices like for example a Pocket PC. By using several active RFID tags (as for example suggested by Brandherm & Schwartz, 2005), it is possible to overlap their signal ranges, which will result in a higher position detection precision (similar to the GPS triangulation described in Section 2.3.2.1).

Similar to active RFID tags are active infrared (IR) beacons. The beacon also constantly emits a unique ID, but instead, it uses the infrared band for communication. The consequence is that a receiver (for example, a mobile device with integrated Infrared Port) can only receive a signal from the IR beacon if it can actually “see” the sender. If there is any object in the direct line of sight between sender and receiver, the signal will be lost. This fact yields both advantages and disadvantages. On the one hand side, it is easy to miss a signal and hence get a wrong positional information (or no information at all), but on the other hand, if an IR beacon signal is received, it is not only possible to estimate the region in which the receiver is located, but also the receivers orientation (as discussed in more detail in 2.3.2.3).

2.3.2.3 Summary

Depending on the application scenario, some techniques may offer a better solution than the others in particular environment settings. In many cases a combination of techniques will be necessary to ensure that the positional information is available at all time and in the desired quality. For example, there is no single positioning technique working efficiently both indoors and outdoors. Hence, in order to allow an application to determine positional information in both scenarios, it will have to employ at least two different positioning techniques. For example, GPS for the outdoor scenario and infrared technology for indoor positioning. A combination of positioning techniques is also necessary if a single technique cannot offer all the necessary data in any given situation. For example, using a GPS for outdoor navigation will offer information on the location but not on the orientation of a device or user. The orientation information can be derived from the GPS trajectory, but this information is only available as long as the user or device is moving. As soon as the movement stops, it is impossible to estimate the orientation. In this case, a combination with an electronic compass may solve the problem. Table 2.1 gives an overview on the different positioning techniques presented in this section.

2.4 Life-Like Character Related Concepts

In this section we will introduce important concepts from the research area of life-like characters. We will discuss typical characteristics of such characters as well as the many different terms used to refer to them within the life-like character community. But before we come to this, we will start the section by taking a look at the origin of the life-like characters.

Life-like characters belong to the research area of Artificial Intelligence (AI), hence we will identify the steps within AI which finally lead to the idea of developing life-like characters. The term Artificial Intelligence was coined during a workshop (McCarthy, Minsky, Rochester, & Shannon, 1955), organized by John McCarthy. However, even though the workshop determines the first use of the term, there have been earlier works which nowadays are associated with the AI research area, like for example the early works of McCulloch and Pitts (1943) and Turing (1950). Since the very early days of AI research, definitions aims of AI varied to a large degree in the view of different researchers. According to Russel and Norvig (2002), there are four main streams of AI definitions:

- Systems acting like humans - the ultimate goal is to program a computer in such a way, that it should be capable of passing the Turing Test[1] (see Turing, 1950) . Main research questions deal with natural language processing, knowledge representation, automatic reasoning and machine learning.
- Systems thinking like humans - the ultimate goal is to program a computer in such a way, that it adapts the problem solving strategies in the human brain. Therefore, it is necessary to understand the complex operations the brain performs. The research area of cognitive

[1]The Turing test is defined as follows: *A computer can think if a person conversing with it over an anonymous communication channel <u>believes</u> that he or she is conversing with a human being*

science is bringing together experimental technologies from psychology and the computer models of AI.

- Systems thinking rationally - the ultimate goal is to realize programming technologies which allow a system to sense, reason and act appropriately. Unlike the approach of imitating human behavior, the idea here is to take actions according to logical rules.
- Systems acting rationally - the ultimate goal is to realize agents striving to achieve the best solution to a given problem by thinking and acting rationally. Even though, rational reasoning is part of this approach, it also deals with situations in which it is impossible to find a provable right decision. In order to deal with these situations, an agent must be able to take a decision even with a certain degree of uncertainty about the results and it should furthermore learn from the observed results of its own actions.

Based on these prototypic definitions of approaches towards AI, the number of specialized subdisciplines within the AI community has been constantly growing over the past decades. Among those disciplines is the field of intelligent agents which we will review in more detail, since the life-like character technology is an offspring thereof.

Russel and Norvig (2002) define agents in general as follows:

> *An agent is everything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators.*

This definition does not limit the term to software agents, but instead may also be applied to humans as actors having eyes, ears and other organs as sensors and hands, legs etc. as actuators. A robotic agent substitutes eyes and ears by sensors like cameras and microphones and may use different types of motors as actuators, while a software agent may take key strokes, databases or network packets as input (sensors) and acts by manipulating data, sending data over the network or displaying information on a screen.

An agent bases its behavior on the information received through its sensors and may potentially take into account an arbitrary long sequence of sensory input when trying to find a solution. The agents behavior is defined by the so-called agent function which maps the aforementioned sequences of sensory data to an agent action. An agents behavior may be completely defined by a table holding all possible sequences of sensory input and the corresponding actions the agent will perform (such a list may be of arbitrary length, since the length of an sensory input sequence may also be unlimited). This is however only true, if the agents behavior does not involve probabilistic decisions. In that case, the table needs to hold probabilities for different actions instead of a single action being linked to a specific sensory input sequence.

The success of an agent is determined by a performance measure based on the goals the agent is trying to achieve. The agent program is designed in such a way, that it strives to maximize the agents performance measure, by taking into account the sensory input sequence and the built-in knowledge of the agent. Early agent implementations featured a static knowledge database. However, modern approaches have introduced the concept of learning through acting to the agent research area, which demands variable knowledge databases. The idea is to allow an agent to perform actions, and to modify its own behavior in the future, in case the results of its actions are analyzed by itself as being suboptimal. The degree of autonomy of an agent is determined by the amount of prior knowledge given to the agent by its designers and its capability of extending its own knowledge-based on experiences it had. The larger the amount of prior knowledge, less autonomous the agent is (according to Russel & Norvig, 2002).

Within the scope of this work, we cannot go into every detail of agent technology, instead we will just highlight one particular aspect which is of high interest to the development of the Migrating Characters.

Agents are put to use in many different task environments which pose different problems to the agent program. A task environment may, for example, be fully observable or only partially observable; it may be static or dynamic and it may feature only a single agent or any number of agents. In the field of software agents, there is a clear distinction between two different classes of agents. The first class realizes so-called static agents whose task environment is limited to the machine on which they are instantiated. As opposed to this, the second class implements mobile agents which are defined as follows:

> *In computer science, a mobile agent is a composition of computer software and data which is able to migrate (move) from one computer to another autonomously and continue its execution on the destination computer*[2].

Mobile agents are characterized by a number of properties. Most importantly, they allow for asynchronous and autonomous operation. They do not need a permanent connection to the machine they were instantiated on. A specific task may for example be performed by an agent migrating to another machine, gathering the desired information, and thereafter coming back to the initial machine to report the results. Therefore, a mobile agent must be capable of adapting in a flexible way to a dynamic environment constituted by the different machines it works on. A mobile agent is capable of deciding autonomously when and where to move next. This movement is realized by allowing the agent to save its own state and transport this saved state to the target host where the agent resumes its execution from the saved state after halting its own execution on the initial host.

In order to realize mobile agents, it is necessary to provide environments which allow for the execution of the machine independent code which constitutes the agent. Hence, the idea is not to evaluate function calls on different machines, but instead, to let the whole agent program as a unit as well as the agents state migrate from one machine to another. This concept of "migrating code" is realized by a number of different agent architectures. Examples of this can be found in Mole (a Java-based mobile system, see also Baumann, Hohl, Rothermel, Strasser, & Theilmann, 2002) and in Grasshopper (the first OMG[3], MASIF[4] and FIPA[5]97- compliant agent platform, see also

[2]Definition found at Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/Mobile_agents

[3]OMG - The Object Management Group, an international consortium founded in 1989 which deals with the definition of standards for system independent, object oriented programming (see also the OMG homepage at http://www.omg.org).

[4]MASIF - The OMG Mobile Agent System Interoperability Facility, a collection of definitions and interfaces that provides an interoperable interface for mobile agent systems.

[5]FIPA - The Foundation for Intelligent Physical Agents, a standard which centers on the notion of agents acting within an agent platform

Perdikeas, Chatzipapadopoulos, Venieris, & Marino, 1999) and we will come back to this concept during the discussion of the Migrating Character concept in section 4.1.2.

In 1985, during his American Association of Artificial Intelligence (AAAI) presidential address, Woody Bledsoe mentioned his continuing dream of building a computer friend (see Bledsoe, 1986). By that time, the most promising sub-discipline within AI to realize this dream was the field of agents. Bledsoe envisioned a machine which he would like to see acting like a human being, a machine which could "understand, act autonomously, think, learn, enjoy, and hate". This "computer person" would be able to read handwritten letters, understand spoken language and translate any utterance in real time from one language to another. It would also be able to recognize faces and help Bledsoe in remembering the names of people passing him by on the street. Bledsoe did not only think of a robot which liked to walk around and play Ping-Pong, but he could also imagine realizing his computer friend as a projected character by means of an overlay included in his glasses. Even though he did not call it a life-like- or virtual character, Bledsoe for the first time described what a life-like character should be capable to do and how it might be realized. Even though not all of the envisioned capabilities of the "computer person" could be realized by means of technologies developed in the research area of agents, there were some promising approaches. Agents at that time were already beginning to show capabilities such as learning and acting autonomously. Some of the agents, developed for multi-agent environments, were even capable of communicating among each other by means of well defined communication interfaces. If we reconsider the definition of the term agent by Russel and Norvig (2002), we should note that human beings themselves may be understood as agents. Hence, the really new problem to be addressed when trying to realize Bledsoe's vision of a "computer person" was that of enhancing a software agent with the capability to efficiently communicate with a human (agent).

Due to the complex nature of human-human communication, the realization of such an human-machine interface demanded the introduction of further fields of research into the development pipeline. Most obviously, the research area of Intelligent User Interfaces is very important in this respect since it deals with the development and evaluation of novel concepts for human-machine interaction. Language understanding as well as language synthesis had been important topics in AI for a long period of time, however human-human communication also involves gestures, facial expressions, and body language in general as well as locomotion. For a robust and most convenient human-machine interface, the "computer person" would have to understand and perform all of the different components of human-human communication. Therefore, computer vision techniques (in order to recognize a human's gestures and facial expressions) as well as computer graphic techniques (in order to realize realistic visual representations of the "computer person", allowing it to move its lips while speaking and to use gestures and facial expressions to support its spoken natural language utterances) needed to be included . By the combination of these different research fields, a new discipline was born, namely that of life-like characters (see Figure 2.2).

One of the earliest characterizations of life-like characters was given by Bates. The author refers to life-like characters as *emotional* (Bates, 1992) and *believable* (Bates, 1994) agents. Bates' approach is based on the "Eliza effect" (see Weizenbaum, 1966), which refers to the fact that people tend to see subtlety, understanding and emotion in an agent as long as the illusion is not actively destroyed by the agent himself. Bates suggests to take a non-traditional view regarding the problem of building intelligent characters. Instead of trying to build characters showing a high degree of intelligence by being especially active and smart, it could be sufficient to try to avoid that

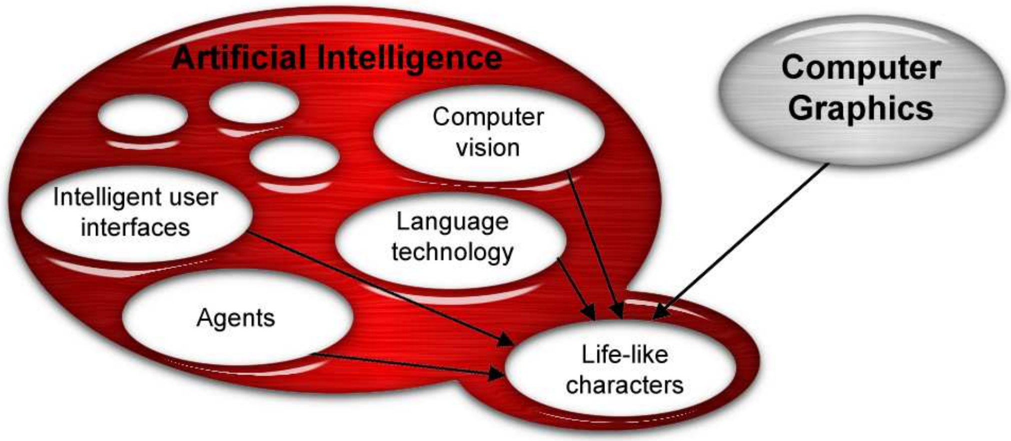


Figure 2.2: The origin of the life-like character research area

those characters appear especially unreal or stupid. To subsume, a life-like character is defined as a virtual creature that computer users are willing to perceive as believable or life-like. Due to the diversity of applications in the life-like character research area, many different terms have been formed by different research groups which describe life-like characters with regard to their special use or abilities within a given scenario.

Anthropomorphic agents, *creatures*, *non-player characters*, *embodied conversational agents* and *synthetic actors* are examples of terms used in the area of life-like characters which are closely related to the specific characters applications (Hayes-Roth & Doyle, 1998; Elliott & Brzezinski, 1998; Petta & Trappl, 1997).

In contrast, Mase (1997) divides life-like characters into three different subgroups based on the different character capabilities:

- **Avatar:** An avatar is constantly controlled by a user and is hence just a visual representation of the user in the virtual world
- **Agent:** An agent may autonomously plan its own movements and hence relieve the user of this task
- **Assistant:** An assistant is an agent which is capable of interpreting plans and following goals in an autonomous way. In addition, the assistant supervises both the user and the virtual environment.

Another example of a term definition in the area of life-like character which is closely related to a specific character competence is that of *embodied conversational agents*, referring to virtual characters which visually embody knowledge about the conversational process (Cassell, 2000).

Even though, all of the above terms and examples refer to life-like characters featuring a visual representation, the concept of "life-likeness" itself is not necessarily limited to this domain. For example, Dautenhahn (1999) discusses life-likeness of robotic agents operating in the physical world while Hayes-Roth and Doyle (1998) point out the difference between *animated* and *animate* characters. The authors refer to *animate characters* as characters featuring all the attributes of life-like characters without having any kind of embodiment. The authors further argue, that the life-likeness of a character is not necessary linked with a life-like appearance or actually any visual appearance at all.

But even within the research groups dealing with animated agents, there is no agreement upon the question of how the visual appearance of a life-like character may influence the perception of the character's life-likeness by the user. The two main streams of character visualizations are *realistic* and *cartoon-style*. The most realistic character visualizations tend to be used in projects trying to create virtual humans (e.g. Kipp, 2003; Gratch et al., 2002; Thalmann, Noser, & Huang, 1997; see also the Virtual Human project described in section 3.3). The majority of cartoon-style characters, on the other hand, are mostly employed in the areas of entertainment and infotainment (e.g. Prendinger, Descamps, & Ishizuka, 2002; Andrè, Rist, Mulken, Klesen, & Baldes, 2000). From a technical point of view, the design of realistic, or even photo-realistic visual character representations, is a challenging task. However, the decision for either type of character visualization also has a psychological component. Users of highly realistic looking characters usually have much higher expectations regarding the intelligent and realistic behavior of the character, as opposed to cartoon-style agents which may make a good impression even with minor inconsistencies.

In a recent publication (Hayes-Roth, 2004), the author goes a step further in her definition of life-likeness of characters by identifying seven qualities of life-likeness :

- **Characters should seem conversational**, capable of engaging other characters or users into conversation in order to exchange thoughts, opinions or feelings. In order to do so, the characters should communicate in a natural way by combining language, gestures, actions and facial expressions.

- **Characters should seem intelligent**, showing a form of expertise that motivates users to engage in the interaction with the character in the first place. In order to seem intelligent, the character does not need to be capable of answering arbitrary questions, but instead it should convey the impression of being an expert in the application area. By channeling the conversation to its own areas of competence and depth, even a character with limited expertise can succeed in conveying the impression of a competent conversation partner.

- **Characters should seem individual**, featuring their own unique and distinctive persona with individual back-stories, personalities and behaviors. Especially in open-ended conversations, users tend to ask characters personal questions. Since these questions are quite predictable, giving a character a personal background and identity will minimize the risk of running into a conversational situation where the character cannot keep up the impression of an intelligent, competent character anymore.

- **Characters should seem social**, allowing them to display self-awareness in the social context. Since many user interactions are inspired by standard forms of social exchange (e.g. praise, insults, flirtation, obscenities, aggression etc.), a character should feature a corpus of understanding knowledge for such communications. However, each individual character should feature its own way of dealing with these situations, depending on the characters individual personality.

- **Characters should seem empathic**, conveying the impression that they actually perceive and express emotion. While each character should feature distinctive emotional dynamics (even though most character engines will base their emotional behavior on the same standard dimensions of emotion like happiness, arousal or dominance) based on different combinations of values on neutral state, provocative events and regression speed, characters should also show a certain degree of awareness of the emotional state of the user or other characters.

- **Characters should seem variable**, by avoiding repetition and by varying all aspects of their behavior unconsciously and incidentally. In order to minimize predictability of behaviors, the characters should make probabilistic choices in what, when and how they act trough all communication channels.

- **Characters should seem coherent**, appearing to be a singular creature, like people. They should be driven by persistent identity and manner and slowly evolving belief structures. Character coherency is determined by all facets of the character personality, including behavior in conversations, typical gestures, facial expressions and body-language.

The author explicitly uses the term "to seem" instead of "to be", indicating that this view is inspired by a combination of an artificial intelligence paradigm with the more artistic endeavors of animators creating comic characters who appear to think and make decisions based on an individual personality. Even though creating a believable comic character is an art of its own, the character's personality only needs to exist in the minds of those who tell the comic stories. Integrating all the above stated qualities of life-likeness in an autonomously communicating life-like character actually requires to model the character's individual personality completely beforehand. In order to measure the success in doing so, Barbara Hayes-Roth suggests to use a slightly modified version of Turings behavioral test for artificial intelligence (Turing, 1950):

> *A character is life-like to the extend that a person interacting with it suspends disbelief that he or she is interacting with a human being.*

Even though this definition is explicitly related to human beings, it may also be applied to characters which resemble "animal-like" behavior instead of "human-like" behavior.

As we have seen in this section, there are almost as many terms used to refer to life-like characters as there are researchers and projects in this area. For reasons of simplicity, we will use the following rule throughout this thesis: whenever referring to a character technology in general, we will use the term *life-like character*, however when reporting on specific projects or character implementations we will use the term suggested by the developers of the particular life-like character.

When reconsidering the definition of a Migrating Character (see section 1.1) we may note that there are parts of the definition very similar to the definition of mobile agents stated above. Mobile agents as well as Migrating Characters have the ability to migrate between different environments and may initiate the migration autonomously. Furthermore both definitions demand that the migration should mean to transfer the state and execution of the agent/character from one environment to another. Similar to mobile agents, a Migrating Character is run inside a software environment allowing to execute the device-independent code which constitutes the Migrating Character. The Migrating Character technology is hence adopting the basic concepts of mobile agents. On the other hand, Migrating Characters are closely related to the research area of life-like characters. The research area of life-like characters and the one of mobile agents are both strongly influenced by the research field of agents and the Migrating Character technology is bringing together concepts and technologies of both disciplines as discussed throughout this chapter and illustrated in figure 2.3.

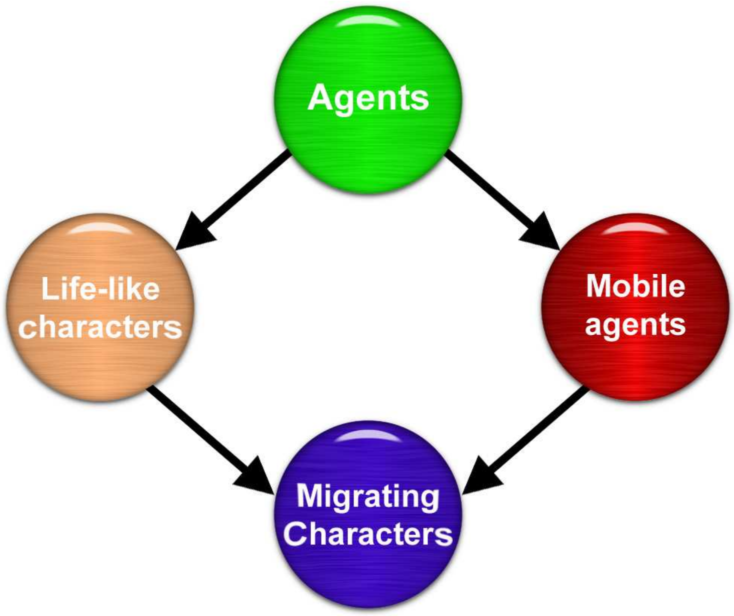


Figure 2.3: Influence of agent technology on Migrating Characters

2.4.1 Social Computing

An intuitive interpretation of the term *social computing* might lead to the assumption that it refers solely to the social aspects of human-computer interaction. However, this is only one way, in which the term is used.

Schuler (1994a) argues that any type of computing application featuring software serving as an intermediary or a focus for a social relation, can be referred to as a social computing application. Examples for such applications, as stated by Schuler, include applications allowing users to communicate via newsgroups, applications defining workplace tasks as well as applications which influence decisions of life and death. This way of understanding and using the term social computing implies also the necessity of including ethical considerations into the discussion. The above stated examples of social computing applications bring up the following exemplary questions with which the research domain of social computing has to deal:

- How do computer based communication channels change the way in which humans communicate and do these communication channels even influence the flow and result of communications (as discussed in Schuler, 1994b)?
- How do computers in the workplace influence or even alter power relations among employees (as discussed in Clement, 1994)?
- How do we assure that a computerized society manages to maintain accountability for the impacts of computing (as discussed in Nissenbaum, 1994)?

Another way of interpreting the term is discussed by Dryer, Eisbach, and Ark (1999). The authors define social computing as a research area dealing with the interplay between persons' social behaviors and their interactions with computing technologies. In their interpretation, the main goal in this research domain is the understanding of relationships among social behaviors and machines, in order to minimize the uncertainty about how humans and machines react. Furthermore, from a technical point of view, the authors wish to facilitate collaboration and support of natural social behaviors by designing information technology systems based on social and behavioral science.

In general, social computing combines research in the areas of interpersonal psychology, computer supported cooperative work and research on social interfaces. Research on social interface theory has a long history and evolves around the fact that humans sometimes respond socially to artifacts. Those artifacts may be either especially designed to foster social responses or more often unexpected social effects will occur. Often, people interact with artifacts as if those artifacts would have human features, like an own personality. The studies described by Nass, Moon, Fogg, Reeves, and Dryer (1995) and Dryer (1999) have shown that machines are sometimes even perceived as entities with individual personalities. Furthermore, the studies have also shown that humans preferred machines with similar personalities to their own over dissimilar ones. There is an ongoing discussion about the question, whether humans actually respond to machines as social actors (as argued for example by Reeves & Nass, 1998) or whether humans can show social responses in human-machine interactions even though the target of such responses is not of any social nature at all (this viewpoint is for example argued by Kiesler & Sproull, 1997). However, it is agreed that certain features of an artifact (e.g. use of natural language, contingent behavior) are encouraging humans to engage in social responses.

Research in the area of anthropomorphic agents is aiming at the maximization of such features within artifacts. Studies have clearly shown that the use of such agents makes interactions with computers more enjoyable (the *persona effect*, as discussed in Lester et al., 1997) and does not result in any negative side-effects (see also Mulken, André, & Müller, 1998). Prendinger and Ishizuka (2004) argue, that anthropomorphic agents promise to have a positive effect on the development of human-computer interaction, since they try to imitate the effective and efficient human-human communication. The authors further suggest, that even though not all interfaces may be improved by adding social components, at least those applications which involve social interaction will benefit from the use of anthropomorphic agents which send and possibly also receive and interpret social cues. Including such social cues into the human-computer interaction will allow for a multi-channel communication between humans and machines. Instead of relying on a single modality, such a communication will integrate a number of parallel information channels, imitating the way human beings communicate (by using gestures and facial expressions together with spoken utterances, as discussed in Section 2.4.2). Prendinger and Ishizuka also state their own characterization of the term social computing as either describing computing which intentionally involves the display of social and affective cues to users, encouraging users to show social responses, or computing which is capable of recognizing affective user states and producing according affective feedback to users.

More recently, researchers are also addressing the problem of interpersonal relationships among life-like characters. As discussed in (Schmitt, 2005), the interpersonal relationships among life-like characters should be reflected in the characters behavior which would result in an increased believability of the characters personality. For example, if two life-like characters in the same scenario are introduced as friends, they should behave accordingly by supporting each other or maybe also by defending one another when (verbally) attacked. Furthermore, these interpersonal relationships should vary over time, depending on the different experiences the characters make. For example, if one character is constantly cheating on another character, one would expect the second character to constantly change its interpersonal relationship to the first character from friend to foe. In some applications it may suffice to pre-script the behavior of the characters and inherently communicate the interpersonal relationship among characters through the script. Schmitt introduces a methodology to model interpersonal relationships among life-like characters which is based on the consistency theories from the research area of social psychology[6]. The author also presents CCNet (Cognitive Consistency Network for Simulating Attitude Changes), a software module which allows to simulate different aspects of group dynamic behaviors among life-like characters. The system provides means to adjust the attitude settings for each character which results in varying choices of verbal and non-verbal behaviors for the character.

With the growing interest in mobile and ubiquitous computing over the past years, the research area around social computing is facing new problems, but at the same time, generating new possibilities. While classical human-computer interaction typically takes place at dedicated locations in front of a stationary computer featuring standard input devices like mouse and keyboard, mobile and ubiquitous computing can basically take place everywhere. Instead of a single user sitting isolated in front of a computer, in mobile and ubiquitous computing researches have to face many different scenarios. For example, multi-user interactions with computers in public locations. Since these applications are no longer fixed to a certain location, it is important to consider social protocols when designing new applications. For example, it is important to consider the social context of a situation, when deciding whether to use speech output or written text in order to deliver a message to the user holding a mobile device.

In a laboratory study, Dryer et al. (1999) have found that the design of mobile or ubiquitous computers may affect responses to social partners. The study investigates the effects on the interaction between two people, one with a computer and one without. The theoretical model behind the

[6]The Cognitive Consistency Theory proposes that people are motivated to change and act consistently with their beliefs, values, and perceptions when there is psychological inconsistency or disagreement between two pieces of information. The conflict between the inconsistent factors produces dissonance. The person begins to doubt previously held rationales, beliefs, or values. These doubts produce uncomfortable feelings and may interfere with the ability to act. The pros and cons of each factor are examined. The resolution of the dissonance occurs when one factor is seen as more attractive than the other. Prior to the resolution of the dissonance, the dilemma between the conflicting factors prevents action. When dissonance is resolved, the person is better able to act in accordance with the more attractive factor because beliefs, values, and perceptions agree with the behavior (Haber, Leach, Schudy, & Sideleau, 1982).

study claims, that the design of mobile and ubiquitous computing devices and applications may not only have an influence on the human behavior (e.g. appeal, disruption, perceiver distraction, user distraction) but also on the social attributions (e.g. agreeableness, extroversion, identification). Both of these factors, according to the theoretical model, have an influence on the interaction outcome (e.g. device satisfaction, productivity, social attraction). The described interrelationships are illustrated in Figure 2.4.

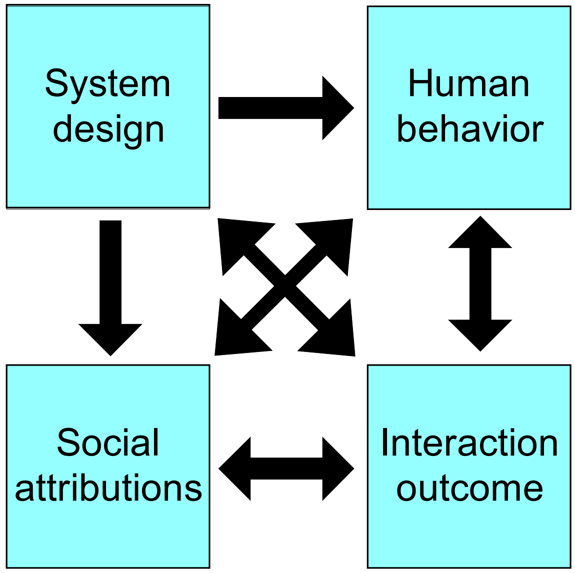


Figure 2.4: Social impact model (according to: Dryer et al., 1999)

Furthermore, due to the physical limitations of mobile devices and applications, new interaction metaphors replacing the old fashioned mouse and keyboard need to be found. The keyboard for example may be replaced by a speech recognition component. Since the displays of mobile devices are usually quite small, replacing written text with synchronized speech is also one approach to overcome a certain limitation of mobile devices. Such a speech-based user interface uses little space while offering high mobility. Since this is a very natural form of communication, it is hence very likely that such an interface would yield social responses of users (as discussed in Dryer et al., 1999).

2.4.2 Gestures, Facial Expressions, Body Language

Communication among humans, as already mentioned in section 2.4.1, is not limited to a single channel of communication. Even though the voice is the most apparent means of communication, quite a lot of information is also delivered via secondary communication channels. These secondary communication channels include gestures and facial expressions, eye-gaze and body language in general. It is interesting, however, that the importance of these secondary communication channels is underestimated by most humans, many are even completely unaware of the fact that they actually use these communication channels at all. While choosing words and sentences is a very conscious process, accompanying gestures and facial expressions seem to be controlled mostly unconsciously. Nevertheless, these secondary communication channels play a very important role in the whole communication process, since they may help to clarify a certain verbal utterance, and they may also help to actually organize a conversation. In face-to-face communications, this may not be evident in the first place. However, in situations where some or all of these secondary information channels are missing, communication among dialog partners is highly effected.

A good example is a dialog via a telephone. In this situation, the communication is limited to a single channel, namely the voice. While it is still possible to perfectly understand one another, certain clues are now missing. For example, in a face-to-face communication, it is possible to tell by the looks of a dialog partner, whether an utterance was ironic or not. On the phone, this additional visual information is missing. Instead, all the information needs to be extracted from the spoken utterance. It is still possible to tell irony by subtle changes in the way a person speaks, but it is harder. Another problem while talking on the phone is the one of turn taking (the back-and-forth interaction needed to have a conversation). While facing a dialog partner it is easy to tell, whether he/she is just taking a short break or is actually waiting for someone else to continue the communication. This is especially important in group communications. However, when talking on the phone (and especially when having a telephone conference with several participants), again, all the information regarding turn taking needs to be extracted from the audio channel.

In (McNeill, 1992), the author identified four different types of speech-associated gestures which ever since have been in the focus of the majority of research projects on the cognitive basis of the gesture-speech relationship:

- **Iconics** are illustrating specific features of the accompanying speech. For example, sketching a circle with one hand while talking about the shape of the Earth.
- **Metaphorics** are a representation of an abstract feature of the current utterance. For example, pointing out the index finger and pushing the hand forward while saying “You must push the button”.
- **Deictics** refer to points in space. Deictic gestures may refer to places as well as to persons and any other entities which are part of the current discourse. For example, pointing at a building and saying “Do you know who lives there?”.
- **Beats** are small formless wave gestures of the hand which usually occur in conjunction with emphasized words, but also when indicating a turn take to another person. For example, by waving a hand briefly up and down to signal a discourse partner to go on speaking.

Many different studies have provided evidence which suggests a close relationship between speech and gestures. For example, McNeill (1992) has found that about three-quarters of all clauses in narrative discourse are accompanied by gestures of one kind or another. Another study (see Alibali & Goldin-Meadow, 1993) indicates, that children start expressing information in gestures even before they start to speak.

Early natural-language dialog systems allowed users to refer to particular entities only by means of linguistic descriptions, often resulting in either ambiguous references or in very complex descriptions (e.g. the ”bright pink flat piece of hippopotamus face shape piece of plastic” in Goodman, 1985). In (Kobsa et al., 1986) the authors discuss the necessity to combine deictic gestures and natural language to simplify and speed up the process of referent identification in natural-language dialog systems. Kobsa et al. present the XTRA system which allows access to expert systems by means of natural language input with combined pointing gestures on objects on a terminal screen.

Cassell et al. (1994) argue that speech and gestures do not always convey the same information, but that the two media rather complement each other. The complementing function of gestures may for example come to use when indicating the respective locations of two objects by the position of two hands, while talking about these objects. The given example in (Kendon, 1994) supports the theory that concepts which are difficult to express in natural language may be conveyed by means of gestures.

In (Kipp, 2003), the author argues that human users are very sensitive and critical regarding bodily behavior of life-like characters. Therefore, a character must act naturally and individually in order to be believable. Kipp especially stresses the fact that each human individual uses distinctive conversational gestures. It is shown that human users are capable of actually recognizing a persons typical conversational gestures when imitated by another person or by a life-like character. It is hence necessary to ensure not only a consistent character appearance, voice and behavior but also a coherent gestural behavior. Furthermore, in applications featuring two or more life-like characters, the conversational gestures of these character should vary notably from each other in order to support the impression of character individuality.

Ever since the early works of Darwin (1898), who was the first to identify and study the importance of facial expressions in human communication, facial expressions have been a very active research area for behavioral scientists. Interestingly, after an initial phase of high interest in the area, in the mid 1950th, scientists generally came to the conclusion that the face should be considered a fruitless source of mostly inaccurate, culture-specific and stereotypical information (this opinion was for example stated in Bruner & Tagiuri, 1954 and is also discussed in Ekman, 1993). The interest in this research area was only revived when both Izard (1971) and Ekman and Friesen (1971) did independently find that there is a high agreement in selecting emotion that fit facial expressions across members of western and eastern literate cultures. However, the most important findings were, that cross-cultural agreement for most emotions was preserved when subjects could use their own words to describe feelings they recognized in expressions (Izard, 1971), and that these findings could even be extended to preliterate cultures which could not have learned those facial expressions and according interpretation from the media (Ekman & Friesen, 1971). The results of Izard and Ekman and Friesen not only inspired other psychologists to further investigate facial expressions and their importance in communication, but they also lead to a raised interest in facial expressions in the research area of computer sciences.

In (Suwa, Sugie, & Fujimora, 1978), the authors present an early attempt to automatically analyze facial expressions. The approach of Suwa, Sugie, and Fujimora is based on the tracking of 20 identified spots in an image sequence. As pointed out by Tian, Kanade, and Cohn (2003), as opposed to human facial expression interpretation which relies on context, body gesture, voice and cultural factors, computer systems need to interpret facial expressions usually regardless of context, culture, gender etc. This so-called higher level knowledge is only applied in emotion analysis in the research area of computer vision.

In computer science, facial expressions have not only been identified as an additional mean for interpreting user input in a natural way, but also for communicating information to the user.

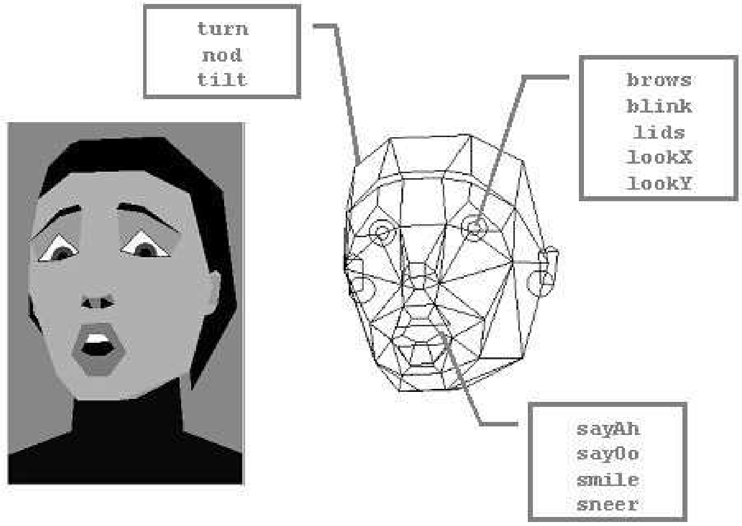


Figure 2.5: A character emotion and corresponding facial muscles (source: Lucena et al., 2002)

The “expressive talking heads” system (see Lucena, 2002) is an example of a virtual character capable of uttering input text by means of speech synthesis which is synchronized with the lip movements of the character as well as the characters emotional expressiveness. The “talking heads” are not simple animations. Instead, they are rendered in real time and they are based on a three-dimensional polygonal mesh with facial muscles that are build by means of vertex grouping. Individual muscles control individual parts of the face, five of the twelve available muscles control eye movements, while three others control the head movements (as indicated in Figure 2.5).

The degree to which humans react on different facial expressions of life-like character has also been in the interest of researchers. Rehm and André (2005) describe two different experiments in which the authors tried to find out, whether the participants would see and correctly interpret subtle clues of lying and deceiving in the facial expressions of a life-like character. In order to realize those subtle clues, the character would imitate human behavior by masking its real emotion by using a fake facial expression. However, these fake facial expressions would not completely cover the facial expression corresponding to the real emotion of the character. Much like with real humans, the resulting facial expression would reveal subtle clues of the hidden emotion below. For example, a true smile will always look different than a smile which is hiding the true emotion. The results of the user studies conducted by Rehm and André indicate that participants were indeed capable of interpreting the facial clues of the character and furthermore, the resulting impression the character made on the participants was also affected.

2.4.3 Attentional Focus Control and Guidance

Telling relevant from irrelevant information in a stream of data coming from all different types of senses is one of the most important capabilities of any living being. The amount of incoming information to the primate’s visual system is for example far higher than the maximum amount of information that can be fully processed by the brain. Many studies have shown that only part of the information is fully processed, while the remaining data are left more or less unprocessed (an exemplary study was conducted by Desimone & Duncan, 1995). Many researchers have tried to understand the way in which the control of attention in primates works. Already over a century ago, James (1890) founded a theory which separates the field of attention control into two subcategories. While James refers to these two fields as *active* and *passive* attention, modern terms are *bottom-up* and *top-down* or *stimulus-driven* and *goal-directed*. The idea behind the distinction is, that in some cases the deployment of attention may depend solely and exclusively on the property of the perceived image (e.g. a sudden, unexpected movement in the periphery of the field of view) while in other situations it may be completely goal-driven (e.g. when searching for a particular object). Even though in theory these two groups of attention control are strictly divided, in reality they seldom occur exclusively. Instead, usually a mixture of both determines the way a visual input is perceived and processed.

There is strong evidence in literature that the distribution of attention can be controlled by an observer intentionally. The classic experiments by Eriksen and Hoffman (1972, 1973), in which the subjects had to identify a letter indicated by a bar marker while attempting to ignore other letters in the display, examined top-down attention control. As a measure of efficiency, the amount of interference caused by the additional letters which were to be ignored was chosen.

The bottom-up control (i.e. capture of attention) has recently been more in the focus of researchers. In (Egeth & Yantis, 1997), the authors introduce two distinguishable categories for stimuli-driven attention control. The first category is called *feature singletons* and refers to stimuli differing substantially in one or more simple visual attributes (e.g. motion, orientation or color). The second category is referred to as *abrupt visual onsets*. While stimuli that belong to the first category capture observers' attention due to their visual distinctiveness from the so-called background (i.e. all remaining objects which are less flashy), stimuli that belong to the second category elicit a reaction of an observer by suddenly appearing in the field of view or by suddenly starting to move and hence contrasting from the background.

Egeth and Yantis also argue that there is a strong interplay between the two different categories of attention capturing. A good example for this interplay is the fact that attentional capture by abrupt onsets can be modified or even prevented by focusing the observers attention somewhere else on the display (as for example in the study of Yantis & Jonides, 1990).

In human-human communication, the coordination of visual attentional focus by means of eye-gaze, head movements and corresponding pointing gestures plays an important role. When talking to each other, humans guide the attentional focus of their communication partners by incorporating gestures (as for example when illustrating the function of a mechanical object by making illustrative hand movements) or by pointing at objects and looking in the direction of the relevant objects.

In an ongoing effort to improve the quality of communication between humans and computers (or machines in general), more and more researchers are trying to incorporate these additional communication modalities into human computer interaction systems. For example, in (Rae, Fislage, & Ritter, 1999) and (Rae, 2000) a project is described, trying to improve the communicative qualities of such a system by improving the interpretation of communicative acts of human users. The authors have developed an active camera system capable of interpreting pointing gestures

Figure 2.6: A binocular camera head, equipped with two RGB cameras, to determine and interpret human pointing gestures (source: Rae & Ritter, visited 2005)

performed by a human user. The prototype, which is based on two RGB color cameras (see Figure 2.6), is not only capable of identifying the user's hands and the performed gestures, but it may also identify an object which is indicated by the pointing gesture of the user. The system emulates human-human interaction by trying to maintain a "common room of shared visual attention" between the system and the user in order to facilitate object referencing and communication in general (e.g. by allowing to use the derived information to control turn-taking, as discussed in Section 2.4.2).

However, the approach described by Rae, Fislage, and Ritter focuses mainly on the interpretation of gestures performed by human users. Equally important, in order to improve the human-machine communication in general, is to enable the additional means of communication also on the machine, or the computer side. In the following section we will discuss, how this goal may be achieved by supporting deictic believability in virtual characters.

2.4.4 Deictic Believability

(Towns, Voerman, Callaway, & Lester, 1998) discusses the effectiveness of life-like animated agents inhabiting knowledge-based learning environments. The authors claim that, due to the strong visual presence of those agents, the whole system promises a significant increase in the students' enjoyment of the learning process. In order to achieve this goal, the authors identify *deictic believability* of life-like animated agents inhabiting artificial worlds as one of the key problems. To allow for deictic believability, a life-like animated agent should be capable of referring to objects in the environment by combining speech, gesture and locomotion. The resulting life-like animated agent should not only be able to disambiguate references to objects in virtual worlds but it is also expected to show increased believability. The authors mention a user study featuring an exemplary life-like animated agent implementation (Cosmo, see also Section 3.3.2) whose results indicate that the agent was capable of producing clear explanations and unique references to virtual objects. In (Lester, Voerman, Towns, & Callaway, 1999) and (Lester, Towns, Callaway, Voerman, & FitzGerald, 2000), the authors generalize their model of deictic believability in life-like animated agents to the field of pedagogical agents. Three criteria are named by the authors which should be met by any animated pedagogical agent:

- **Lack of ambiguity:** References performed by a life-like animated agent in a virtual learning environment should be unambiguous in order not to confuse the students. This is especially true in learning environments where false interpretation of agent references might lead to misunderstandings and failure of students.
- **Immersivity:** Much like humans are usually immersed in physical environments while performing spatial references to physical objects, life-like animated agents should immerse into the virtual world by moving within scenes and towards objects while referring to those objects.
- **Pedagogical soundness:** The deictic gestures of life-like animated agents in virtual learning environments should support the ongoing advisory discourse and hence also support the central pedagogical intent as a whole.

While the aforementioned user study by Towns, Voerman, Callaway, and Lester has shown a general improvement in the quality of virtual object references when performed by a deictic believable life-like animated agent, the question whether deictic capabilities improve an agents believability is harder to answer. As Allbeck and Badler (2001) mention, the term *believable*, when used in the *embodied agents* research community, usually refers to an agent whose behaviors are created with an attempt to conform to our nominal expectations. The term believable on its own is already awkward to define, it is generally in a sense like "if something is believable, it is accepted as being real". However, this definition demands a clear understanding of what is *real*. Since a life-like animated agent cannot resemble all the aspects of a real person, believability of those agents is often defined by the agent's actions and communications which should be as close as possible to a real persons according behaviors. According to this definition of a life-like agent's believability, it is of utmost importance to support as many manifestations of nonverbal communication in an agent as possible, since these play such an important role in human-human communication (as already discussed in section 2.4.2).

As pointed out in (Khullar & Badler, 2001), eye behaviors (or gaze) also play a very important role in human-human communication in general, and particularly when referring to physical objects. When referring to a particular object in the environment, the referent's viewing direction and angle are important clues for the communication partners about the location of the object the speaker is referring to. A life-like agent pointing in the direction of an object while verbally referring to it, should also imitate human behaviors by looking in the corresponding direction. However, eye movements and gaze are also generally important in human-human communication. While communicating, speakers use glances to emphasize particular phrases or words, while listeners signal attention or a deepened interest in a particular topic. In addition, gaze is important in terms of regulating the flow of communication (turn taking, as discussed in section 2.4.2).

An important question however is, whether users of a life-like character system are actually willing to communicate with these characters in a very human-like way. Up to which degree are users willing to accept a life-like character as a coequal communication partner. In human-human communication there are many social protocols involved. Most humans participating in communications try to follow those protocols in order to avoid to appear disrespectful. It is however not clear, whether human users of life-like character systems are willing to accept the same social protocols as in human-human communication.

The degree of acceptance of social protocols when interacting with life-like agents is an open question. However, researchers are starting to investigate in that direction. In (Prendinger, Ma, Yingzi, Nakasone, & Ishizuka, 2005) for example, the authors report on a user study which has been conducted in order to find out, whether users follow the gestures performed by a simple virtual character and whether they do this in a similar way as in human-human communication. The results of the experiment are based on information derived from an eye tracker device which allowed to calculate the users gaze direction at any given time. The bottom line of the experiment is, that users did not only interpret the gestures of the life-like character correctly and focused their attention on the according virtual objects, but that users also kept up a behavior known from human-human communication. Instead of keeping their focus on the referred object after the reference was performed by the character, the user's gaze went back to the character. It actually even went back to the character's face, which is what one would expect of a polite listener in a human-human communication.

2.5 Synopsis

In this chapter we have reviewed several concepts and factors relevant to the development of the Migrating Character technology. After a brief discussion of the term "instrumented environment", we subdivided the remainder of this chapter into three main sections, each focusing at a different research area. While the first section dealt with basic concepts related to object referencing methods, the second section focused on the research areas of mobile and ubiquitous computing. Finally the third section reviewed relevant concepts from the area of life-like characters. The Migrating Characters described in this work do combine elements from each of the three different research areas which were the main focus of this chapter. In this section we will discuss how different aspects of those research areas, which in combination will form the basis of the Migrating Character concept, may complement one another.

In contrast to "standard" life-like characters, the Migrating Characters enter a completely new realm. By leaving the single display, fixed location environment, the Migrating Characters face new opportunities but also have to deal with problems unknown to prior life-like character implementations. While spatial concepts may also play a role in virtual worlds (especially in three dimensional virtual worlds), the importance of these concepts in mobile and ubiquitous computing applications is far more obvious. Even though not every mobile computing application or service automatically deals with spatial concepts and problems, the fraction of mobile applications taking into account spatial information is rapidly increasing. While early mobile services mainly offered the benefit of being available at arbitrary locations, like for example with the advent of mobile phone technology, nowadays mobile applications are often offering additional, location-based services. A very prominent example is car navigation systems. Even though these two examples of mobile computing services belong to different categories of mobile computing, they share a common characteristic, namely the use of unreliable communication channels and data sources. It is a common peculiarity of all mobile applications that they have to deal with situations in which certain access to specific data resources is limited (i.e. low bandwidth of wireless communication channels, low precision of positional data) or often completely unavailable. The ways in which mobile services deal with these problems are very diverse. While mobile phones, when losing connection to the providers network, will simply inform the user about this fact, car navigation systems try to compensate the risk of running out of data by incorporating several redundant information sources (i.e. the data from the global positioning system is supported by data derived from sensors mounted in the car). If one data source is lost, the system may continue to work (even though the precision of the derived data may be affected).

Unreliability of data sources and communication channels is however not the only problem which modern mobile applications have to deal with. The mobility aspect of these applications demands ever decreasing device sizes and weights. While the computational power of such small, embedded devices is constantly improving, the size of the integrated displays in mobile devices cannot be increased. Even though the quality and resolution of mobile devices' displays is also improving, the visual cognitive system of human users constitutes a natural barrier for further improvement in this area. Hence, it is necessary to find an effective means for using the small screen space in the most effective way. This is especially true for devices which do use the display not only for presenting information to the user but also to support human computer interactions by means of integrated touch screens. In this case, the small screen space has to be shared between presentation space and interaction space. It is hence necessary to find new interaction metaphors for mobile applications.

Regardless of the chosen interaction and presentation method for a mobile service, there are certain types of information and presentation media which may not be presented in an appropriate way on the small screen of a mobile device. Very detailed, high quality images and video clips for example may demand the use of a larger, high resolution screen in order not to risk to loose important details. While large scale images may be transformed into a sequence of images showing relevant parts of the image (see also section 6.1.7) in order to fit on the mobile device's screen without sacrificing the level of detail in the images, video clips need to be seen as a whole.

The Migrating Characters try to deal with the limited screen size of mobile devices and the resulting problems by means of different adaptation strategies. A Migrating Character is used as a general interaction metaphor, however its behavior changes due to the actual user context. Instead of solely relying on the mobile device, the Migrating Characters incorporate available technology in the environment. The idea is to maximize the input and output quality of a mobile computing system by automatically determining available, additional input and output devices in the vicinity of the user. These devices should be used whenever the quality of both input and output would be lower if only the mobile device would be used. The combination of mobile and stationary devices in a flexible way demands a consistent user interface in order not to confuse the users while interacting with the system. The Migrating Characters, being capable of transferring from a mobile device onto a stationary device in the environment, offer such a consistent interface metaphor. The Migrating Character is used as a unique interaction interface, but also as an anchor point for the users while presentations are spanning over mobile and stationary devices. By jumping from one device to another, the Migrating Character guides the users attention towards different devices during multi-device presentations. Furthermore, since mobile applications are interacting with the physical world, the Migrating Characters' attentional focus guidance capability may also be expanded in order to guide a user's attention towards a physical object. For example, by moving the character close to the object and letting it perform according gestures and verbal utterances to clarify the reference to a particular object (as described in detail in Section 6.3).

3 Previous and Related Work

While the preceding chapter summarized basic concepts from the research areas of object referencing, mobile and ubiquitous computing and life-like characters related to the development of the Migrating Character technology, this chapter will give an overview on the scientific state of the art in these research areas by reviewing a selection of prominent and/or closely related systems. Unlike the last chapter, this chapter is subdivided into six sections. The projects described in each of the sections were chosen in order to provide a background for the work presented in the later chapters.

We will start by reviewing relevant projects from the area of life-like characters. This section is then followed by an overview on projects which address localized information presentation based on mobile devices. We will then continue by introducing a number of projects which focus on the development of life-like characters. These characters inhabit virtual worlds and are capable of guiding users through these virtual worlds. The next section deals with projects which focus on life-like characters used in mixed reality applications, meaning a combination of virtual and physical environment. In the last two sections we will first review several different approaches towards a combined use of mobile and/or stationary devices and then we will conclude our research overview by looking at projects which support life-like character technology in mobile applications.

Each section will be followed by a brief summarization of the important aspects of the projects discussed in the section (with respect to the development of the Migrating Character technology).

After reviewing the different systems and prototypes, we will conclude the chapter with a discussion of the different projects' advantages and shortcomings as well as possible aspects which might influence the realization of the Migrating Characters.

3.1 Projects Aiming at Different Aspects of Life-Like Characters

In this section we will review a selection of projects related to life-like characters. The selection of projects was taken to give a short overview on the state of the art in life-like character technology. Each of the projects introduced in this section focuses on a different aspect of life-like character technology. While the CrossTalk project's main focus is on the communicative skills among life-like characters, the Virtual Human project aims at the realization of real-time interactive communication between human users and life-like characters. Unlike the first two projects, the Galatea project's main goal is to create photo-realistic appearances for life-like characters.

3.1.1 CrossTalk

CrossTalk (Baldes et al., 2002) is an interactive installation developed at the German Research Center for Artificial Intelligence (DFKI) based on the Avatar Arena approach(see also Rist & Schmitt, 2002) developed within the scope of the MagicSter project[1]. In Avatar Arena, simulated negotiation dialogs with affective, embodied conversational characters (embedded in a social context) are staged. The basic idea is to allow users to send their delegates (avatars) to a virtual environment where the avatars negotiate on behalf of the users. Crosstalk is one example of an avatar installation allowing users to manipulate the avatars goals and attitudes and to watch the resulting dialogs among the avatars. It features three virtual agents performing simulated dialogs to present product information. While the first agent, Cyberella, introduces and explains the idea of simulated dialogs among virtual agents, the two other agents, namely Tina and Ritchie, engage in a car sales dialog (Rist et al., 2002). The CrossTalk project proposes a shift from single agent settings to performances given by a team of agents as a new presentation method. The main idea was inspired by the evolution of TV commercials over the past 40 years. While in the early TV commercials a single salesperson would enumerate the positive features of a specific product, nowadays we see commercials showing dialogs among several actors or even small episodes. The product information in these commercials is transported in a more subtle way, by being integrated in a whole story. The repertoire of presentation strategies in the field of life-like characters is largely enriched by this idea.

The CrossTalk interactive installation consists of a central control panel with a touch screen and two additional screens to display the agents (the Cyberella agent is shown on one screen, while Tina and Ritchie are shown together on the second screen). In addition, each of these two additional screens is combined with an audio system used to render the verbal utterances of the agents shown on the corresponding screen. The spatial layout of the installation is that of a triangle in which on each corner there is one screen. These screens are all oriented towards the user which is standing in front of the interactive touch screen. In addition, the area in front of the interactive touch screen is monitored in order to detect approaching users and to automatically react to their presence. Each of the three main elements of the installation (i.e. the Cyberella agent on one screen, Tina and Ritchie on the second one and the interactive user interface on the third screen) are run on separate computers. The communication between these elements is handled by an information router based on TCP/IP sockets. The CrossTalk interactive installation is depicted in Figure 3.1.

[1]The MagicSter project homepage: http://www.ltg.ed.ac.uk/magicster/

Figure 3.1: The CrossTalk installation as shown at the CeBIT computer fair in 2002 (source: Rist et al., 2002)

The CrossTalk interactive installation utilizes the Inhabited Market Place (IMP, see Andrè et al., 2000) as an exemplary application scenario. IMP is a virtual market place that employs presentation teams to convey information about products. The user, observing the staged dialogs between the different members of these presentation teams, learns about the features of the product throughout the dialog. However, these dialogs are not scripted but adaptable. The CrossTalk installation allows users to assign different roles and attitudes to the different members of a presentation team before the actual dialog starts. In this way, a variety of different sales dialogs can be generated for one and the same product.

3.1.2 Virtual Human

The *Virtual Human* project (Göbel et al., 2004; Pfleger & Alexandersson, 2004), a joint project lead by several main project partners (the German Research Center for Artificial Intelligence (DFKI), the Fraunhofer-Institut für Medien-Kommunikation (Fraunhofer IMK), the Fraunhofer-Institut für Graphische Datenverarbeitung (Fraunhofer IGD) and the Charamel GmbH), aims at the realization of "realistic" anthropomorphic interaction agents, capable of engaging in real-time human-like affective communications with users. Instead of using pre-scripted dialogs for multi-party conversations, the agents in the Virtual Human project are driven by a behavior engine capable of coordinating the exchange of speaking turns in order to avoid incomprehensible speech due to overlapping utterances. The project combines state-of-the-art computer graphics technology with a multimodal dialog component which was developed within the scope of a previous project at the DFKI (SmartKom, see Wahlster, 2003 and also Subsection 3.5.6).

From a technical point of view, the Virtual Human platform is subdivided into a number of modules which communicate with each other by means of well defined interfaces. This modular approach allows for different setups, allowing to combine different subsets of the available modules in order to fulfill a specific application task. An overview of the main modules forming the Virtual Human platform is shown in Figure 3.2.

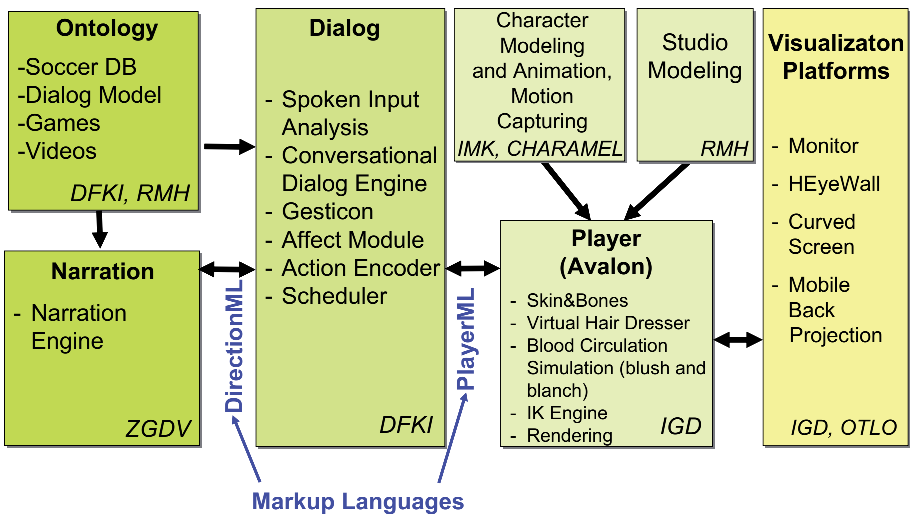


Figure 3.2: The Virtual Human Platform - Component-based Architecture (source: Wahlster, 2005)

The authoring module contains a comprehensive authoring environment including several editors allowing developers to configure application scenarios, define interaction metaphors among virtual humans and between users and the Virtual Human system, and to create stories and agents.

The main function of the narration module is to control the narrative and didactic progress within a session. Based on concurrently executed declarative story and learning models, different scenes, the elementary portions of the variable story line, are chosen. In this way, during run-time story creation, scenes are chosen out of a pool of available scenes.

The dialog module is responsible for the generation of action scripts to be interpreted by the player module later on. The dialog generation is subdivided into three steps. After the dialog is initialized with a very common representation in the first step, the second step adds additional information to the dialog description (e.g. timing for lip-synchronization, emotion). The final step enhances the dialog script by using fitting animations like facial expressions, gestures and other animations. The resulting dialog description is forwarded to the next module, the player.

The player module (namely the Avalon player, a component-based virtual and augmented reality system) is extending the concepts of VRML[2]/X3D[3] (Behr, Dähne, & Roth, 2004). The virtual world in Avalon is described by a scene graph which defines geometric and graphical properties of the scene as well as the its behavior. For the Virtual Human platform, a new functionality was added to the Avalon player, allowing it to interpret scripts generated by the Virtual Human dialog module.

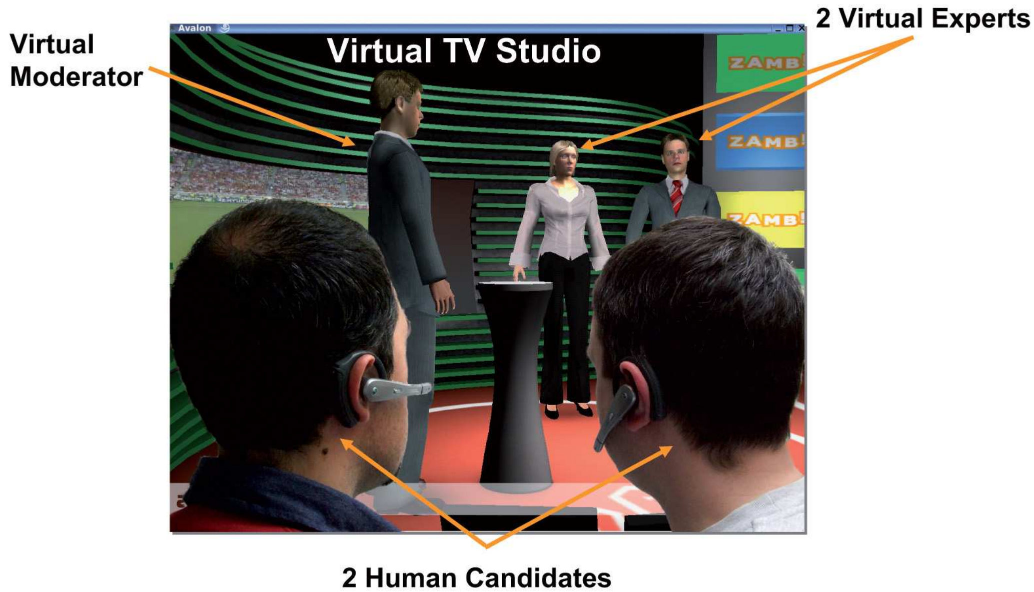


Figure 3.3: The Virtual Human Demonstrator as shown at the CeBIT computer fair in 2006 (source: Wahlster, 2005)

For communication purposes between different modules within the Virtual Human platform, two different markup languages were defined. While the Direction Markup Language (DML) represents an XML-based script-language to initialize a scene (i.e. it describes all static scene items, for example the background or the position and orientation of the virtual characters), the Player Markup Language (PML), which is also XML-based, allows to describe the three-dimensional space containing objects and virtual characters and also possible actions for these objects (e.g. interaction, movement). Within the Virtual Human platform, DML is used for the communication between the narration module and the dialog module, while PML is a mean to exchange data between the dialog module and the player module.

As mentioned above, the modular setup within the Virtual Human platform allows for different demonstrator setups resulting in different “levels of intelligence” (see Göbel et al., 2004). While the simplest level consists of predefined scene dialogs (created with the authoring environment mentioned above) directly sent to the player, the most elaborate intelligence level allows for bidirectional communication between the different Virtual Human modules, resulting in a variable story taking into account events caused by user interactions.

The Virtual Human demonstrator shown at the CeBIT 2006 computer fair (see Figure 3.3) three

[2]the Virtual Reality Markup Language

[3]an open standards XML-enabled 3D file format, successor of VRML

virtual humans were interacting with up to two human participants in a tv-like quiz show. The team of virtual humans consisted of a moderator and two football experts. During the presentation, the participants were shown video clips of famous football scenes which were stopped at the dramatic climax of the scene. The participants had to guess how the scene would end (e.g. the player will score or miss the goal) and could ask the experts about their opinion. Each of the virtual humans could be directly addressed by the participants and the virtual humans themselves also addressed each other as well as the participants directly.

3.1.3 Galatea

The *Galatea* open-source software for the development of Anthropomorphic Spoken Dialog Agents (ASDA) is being developed by a consortium of universities and research institutes lead by the Japan Advanced Institute of Science and Technology. The main goal within the project is to develop a software toolkit allowing to simplify the development of human-like spoken dialog agents (see Kawamoto et al., 2002).

The Galatea software toolkit combines several different modules including a speech recognizer, a facial image synthesizer and a dialog controller. Each module is modeled as a virtual machine with a simple interface and is connected to a communication manager, or a broker.

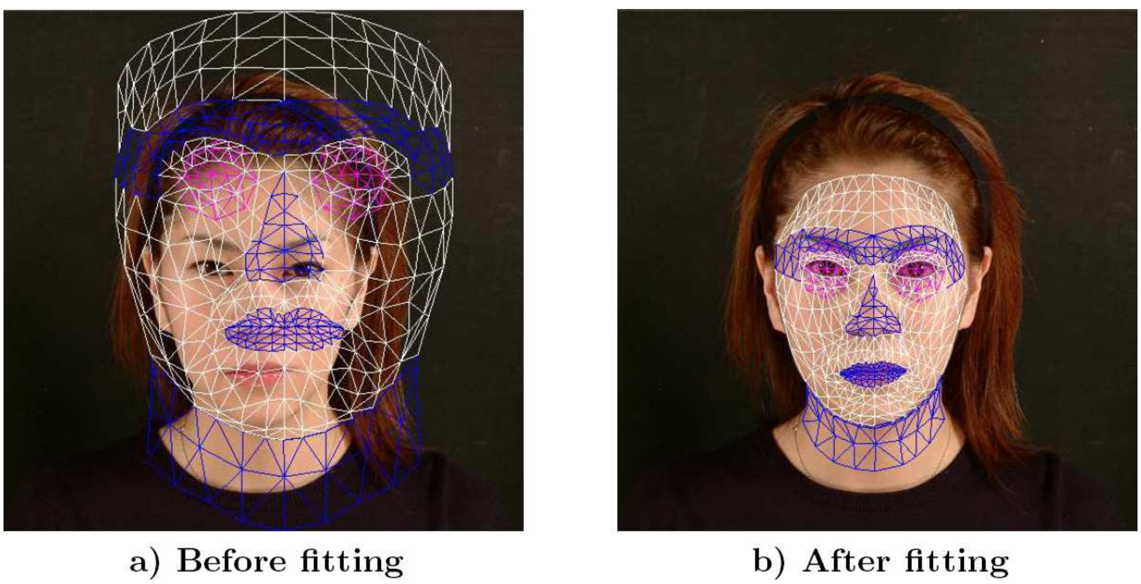


Figure 3.4: Fitting a facial model on a photograph with Galatea (source: Kawamoto et al., 2004)

Using these modules in combination allows to realize ASDAs which are capable of performing face-to-face communications with users similar to natural human-human communications. To achieve this goal of a very intuitive human-computer interaction, the ASDA must imitate human behavior during conversations which consists of a combination of spoken utterances, gestures and facial expressions.

Galatea allows developers to build ASDAs with different visual appearances and different voices in a very convenient way. During the development phase of an ASDA, several different tools come to use. A highly customizable text-to-speech synthesizer allows developers to individualize the voice of an individual ASDA by modifying corresponding features in the synthesizer. To allow for individual visual appearances of the ASDA, Galatea makes use of an innovative facial animation synthesis. The synthesis is based on a single photograph of a real person. This frontal photograph is then opened in a special editor, which helps the developer to customize a generic face model (i.e. a 3D mash representing a generic human face) in order to fit the individual features of the person whose photograph is being used (this process is briefly illustrated in Figure 3.4 and is described in detail in Kawamoto et al., 2004).

Figure 3.5: Examples of typical expressions rendered with Galatea (source: Kawamoto et al., 2004)

In a next step, using the mouth-editing-tool, 17 controlling parameters corresponding to 17 lip parts are modesfied in order to fit the photograph. Since all mouth shapes for viseme[4] in English

[4]A viseme is a generic facial image that can be used to describe a particular sound. A viseme is the visual equivalent of a phoneme or unit of sound in spoken language. Using visemes, the hearing-impaired can view sounds visually - effectively, “lip-reading” the entire human face (source of this definition: http://whatis.com).

and Japanese are already predefined in the generic model, the new ASDA is automatically capable of performing spoken utterances with according, realistic lip movements (to achieve synchronous spoken utterances and lip movements, a real-time synchronizing module is used).

Facial expressions are generated by combining the so-called action units (AU). The face is divided into 44 AUs; each corresponds to a facial muscle movement. These AUs control the basic movement of the face (e.g. control the upper lip raiser or the inner brow raiser). In Figure 3.5, four typical facial expressions generated by Galatea are depicted (from upper left to lower right: happiness, sadness, anger and fear).

3.1.4 Summary

Even though the three projects introduced in this section all have a very different approach towards life-like character technology, they all share a common goal: to improve the degree of realism of life-like characters which consequently will lead to an increased acceptance of such systems by the human users. By imitating the communication among humans with life-like characters the communication between computers and users becomes more natural and intuitive. In order to convey the impression of life-likeness, it is important to design the behavior of the character as human-like as possible. When communicating among each other or with human users, life-like characters should not only rely on spoken utterances but instead, they should utilize all the secondary communication channels like gestures and gaze, which come to use in human-human communication, as well (see also Section 2.4.2). Three preconditions need to be met to allow a life-like character to successfully imitate the complex human-human communication. First of all it must be able to recognize and interpret natural language input as well as gestures and facial expressions of users (and possibly also other characters). Secondly, it must be capable of generating appropriate answers, or utterances in general, and then it should also be capable of determining according gestures and facial expressions to support the natural language utterance. However, the first two conditions are useless if a life-like character does not feature an appropriate visual representation allowing it to realize the planned gestures and facial expressions in such a way that a human user can successfully interpret the communicative acts of the character.

3.2 Mobile Applications Offering Localized Services

In this section we will give an overview on projects involving mobile computing. We will, however, focus on projects using localized information since those are especially relevant for the development of the Migrating Character technology. Positional information can be useful in many different ways in mobile applications. It may be used in instrumented environments to locate specific users of the system and to offer comfort functions to the users, for example, by redirecting a phone call to a phone physically close to the users location (an example is for this is the Active Badge project, see Section 3.2.1). Furthermore, the user's positional data may also be used in a navigation context, allowing a system to automatically direct a user towards a specific location. Using different positioning techniques, allows for both indoor navigation (see for example the IRREAL project described in Section 3.2.4.1) and outdoor navigation (the DeepMap system for example uses the GPS system to locate users in outdoor scenarios, see also Section 3.2.6). Finally, location-based services may also feature information presentation which is directly linked to the user's position or context. The ARREAL system (see Section 3.2.4.2), for example, takes into account the user's actual position and walking speed when automatically annotating a map showing the surrounding of the user, while the Sotto Voce project (reviewed in Section 3.2.7) offers information on specific objects in a historical house, depending on the users position. As opposed to all the other projects reviewed in this section, the Sotto Voce project does not rely on a hardware driven solution to determine a users position. Instead, it expects the users of the system to indicate when they move from one room to another or whenever they turn towards a different wall in the room.

3.2.1 Active Badge

The Active Badge location system (see Want, Hopper, Falcao, & Gibbons, 1992) was developed with the goal to simplify the integration of telephone systems with computer systems. While conventional systems make use of "pager systems" as a solution for personnel location, the Active Badge system offers a completely new approach. The aforementioned "pager systems" make use of a central facility which sends out signals to individual receiver units (the so-called "beepers") which react to this signal by producing an audible signal. In addition, the beeper may also display a number which the paged person should call back. In order for the system to produce the desired result (i.e. the transfer of information between the initiator and the receiver of the signal), the user is expected to take appropriate actions by himself after receiving a signal.

The Active Badge system addresses several weaknesses of the "pager system" approach:

- Whenever a signal is sent to a user and the user does not respond, the "pager system" cannot determine the reason for this. Possible causes are: the users beeper may be inactive or out of reach, the user may have missed the audible signal
- "Pager systems" cannot determine the physically closest user to fulfill a certain task (since they do not know the locations of the users)
- A beeper will always produce an audible signal, regardless of the context users are in at the moment the signal is received. This may be a disturbance for the users and may force users to turn off their pagers in certain situations

The Active Badge idea is to equip users with a small device which is traceable within a certain scope. Instead of contacting the user and waiting for a response over conventional communication channels (e.g. telephone), the Active Badge networked system uses the positional information of each user to determine the communication facility closest to the user's actual position. The communication with a particular user is then automatically initiated using the determined communication facility. Since the system is aware of the locations of all users, it may adapt to different situations. It may, for example, automatically determine which user (with a certain expertise necessary to fulfill a certain task) is closest to a location where the user's expertise is needed. Instead of contacting all the experts known within the system, it will only contact the closest one. The system may also determine situations in which a user should not be disturbed (e.g. a user may define within the system that he/she does not want to be disturbed while being in the office of his/her boss) by analyzing the location of the particular user and also taking into account the locations of other users close by.

Several different Active Badge devices were developed between 1989 and 1992 at the Olivetti Research Ltd (see Figure 3.2.1 for a selection of four different Active Badges). While the general design goal (a small, lightweight device) was kept throughout the development, the technical setup changed from a simple infrared transmitter sending a unidirectional 5 bit code to a complex infrared transmitter and receiver integrating a 87C751 microprocessor allowing the device to send a 48 bit code and supporting bidirectional communication. Each of these senders emits a unique code identifying the user every 15 seconds. The code emitted by the Active Badges is received by a network of infrared receivers, the so-called Badge Sensors (see figure 3.6). These sensors are spread throughout the environment allowing for a good coverage in order to receive the codes of the individual Active Badges. Up to 128 of these badge Sensors can be connected to a single serial port of a standard workstation. The sensors are powered through this network. These networks of up to 128 sensors are then combined over an Ethernet connection allowing to build networks with large quantities of sensors.

Figure 3.6: left: Four generations of active badges, right: the networked sensor (source: http://www.uk.research.att.com/ab.html)

Within the Active Badge project, several different applications using the Active badge technology have been developed, such as telephone routing using conventional telephone networks, automatic email rerouting to different workstations, and print-job rerouting to the closest printer. The many different applications realized with Active Badges are discussed in detail by Want & Hopper, 1992.

3.2.2 ParcTab

The ParcTab system was developed as a prototype at the Xerox Parc research lab. The main goal of the project was to evaluate different approaches for the use of mobile devices in office scenarios (see Want et al., 1995; Want et al., 1995). The ParcTab system is a combination of Palmtop computers which are connected via a wireless infrared connection to applications running on different workstations within the office environment. The ParcTab Palmtop computer and a infrared transceiver that allow for bidirectional communication and feature network functionality are shown in Figure 3.7.

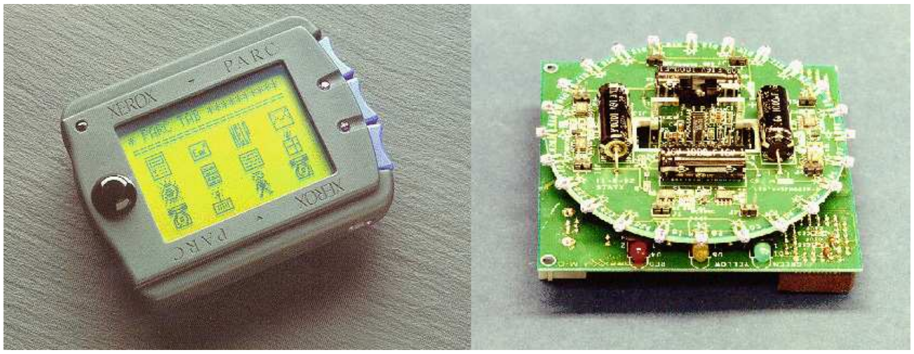


Figure 3.7: left: The ParcTab mobile computer, right: The ParcTab infrared sensor (source: http://www.ubiq.com/parctab/)

The ParcTab Palmtop computer was designed in such a way that the users may easily take the device with them wherever they go. The size, weight and features of the device allow for spontaneous usage of computers in any given location and at any time. The Palmtop device does not feature a power switch, since it is automatically turned on whenever the user starts to interact with the device and it also turns of autonomously after a certain period of inactivity. The device has a constant network connection allowing it to communicate with other devices and services within the same network. The location of each ParcTab Palmtop computer is available within the system at all times and is derived from the location of the infrared transceivers communicating with the mobile devices.

The system offers context information, for example, location, time, and other users close to the user in focus. This context information is used to adapt the behavior of the many different applications which were developed within the scope of the ParcTab project. The applications were subdivided into classes of applications sharing certain attributes. For example, the *contextual reminders* remind a user of certain events which take into account not only the time at which the event occurs but also the location of the user and the presence of others around the user. The *Tabdraw* application allows users to drop virtual post-it notes on the ParcTab Palmtop devices of other users. Further applications are an email client, a scheduler and an application which allows users to determine the location of other users of the system within a building.

3.2.3 DoC - Lancaster University - GUIDE

The GUIDE system was developed at the Department of Computing (DoC) at the Lancaster University to provide *an intelligent electronic tourist guide* (as described in detail in Cheverst, K., Davies, N., Mitchell, K., Friday, A., & Efstratiou, C., 2000) to the city of Lancaster. The GUIDE system combines several different techniques (context awareness, personal computing technologies, wireless communications and adaptive hypermedia) to support the information and navigation needs of a visitor to the city of Lancaster. The systems design is based on a distributed cellular architecture(see Figure 3.8). The information required for both navigation and information issues is broadcast by cell-based stations to mobile devices. These broadcasts may be scheduled or in response to user requests (as discussed in Davies, N., Mitchell, K., Cheverst, K., & Blair, G., 1998). The information broadcast is based on the location and the user preferences. Information for a given geographic area is broadcast by specific base-stations mounted in the region. The system supports a *back-channel* via both fixed and wireless links, enabling interactive services and time-critical information broadcast. The mobile devices run a customized web-browser to present the information to the user. The Fujitsu TeamPad 7600, featuring a transflective screen to improve readability in direct sunlight, was chosen as the end-system. While driving a wireless network card, the running time of these devices was approximately two hours(which was considered suitable for evaluation purposes but not enough for serious use of the system).

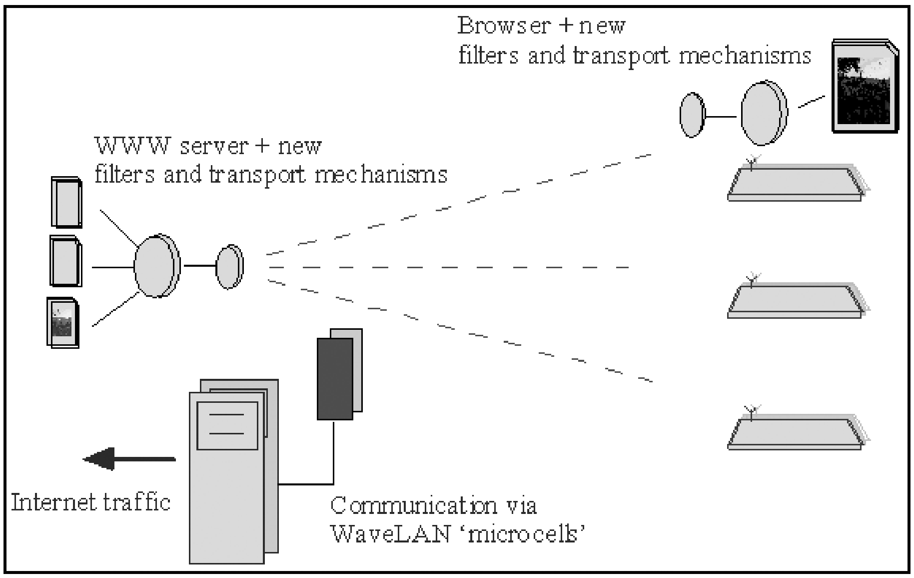


Figure 3.8: The GUIDE Architecture (source: Davies, N. et al., 1998)

The user interface allows a user to choose among several different tours, to navigate to a specific geographic region or to get detailed information on a region and its nearby attractions. Available options vary due to the user's context and preferences. Tours also vary according to opening times of sights to visit. The GUIDE system has been evaluated in two phases:

- an expert walkthrough
- a field trial

The field trial involved sixty people who volunteered to use the system. Results showed that the overall opinion was that the context-aware navigation and information retrieval were both useful and reassuring (results are discussed in Cheverst, K., Davies, N., Mitchell, K., & Smith, P., 2000).

3.2.4 REAL

The REAL project is part of the "Sonderforschungsbereich 378 - Ressourcenadaptive Kognitive Prozesse" (see Wahlster & Tack, 1998). This special research group deals with resource adaptation in cognitive processes. The central question to be answered in the course of the REAL research project is: "How may a mobile person navigation system adapt to the recognized technical and cognitive resource restrictions of its user, by altering the presentation of way descriptions?" (this question is discussed in Baus, Krüger, & Butz, visited 2005).

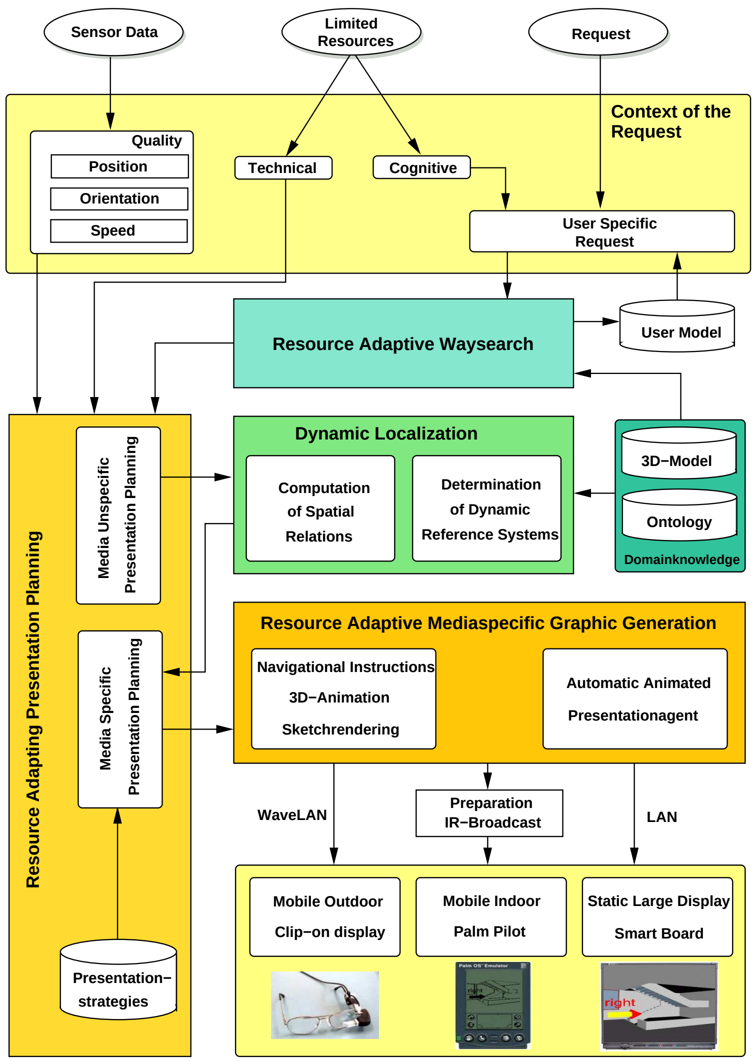


Figure 3.9: System architecture of the hybrid person navigation aid (source: Baus et al., visited 2005)

Within the scope of the REAL project, a resource-adaptive mobile navigation system is developed (described in detail in Wahlster, Baus, Kray, & Krüger, 2001). The techniques developed in the course of this project are implemented and evaluated in two different scenarios. While the IRREAL sub-project (Butz, Baus, & Krüger, 2000) is used for indoor (building) navigation, the

ARREAL sub-project (Kruppa, 2001) is used for outdoor navigation. Both sub-projects are combined to build a hybrid(both indoor and outdoor) person navigation system. Figure 3.9 shows the integration of both sub-projects into the combined navigation system. Important aspects of the different sub-projects (belonging to the REAL project) are discussed in the following subsections. In addition, the REAL system offers the opportunity to use a stationary information system in order to prepare the navigation task. The stationary component is explained in Section 3.5.5.

3.2.4.1 IRREAL

The IRREAL scenario implements a building navigation system based on hand-held PDAs and a number of strong infrared transmitters. The PDAs are standard devices such as 3Com Palm Pilots or the Handspring Visor are used as display units. The PDAs make use of their built-in IrDA interface to receive up-to-date information and directions transmitted via the infrared transmitters. Figure 3.10 shows an infrared transmitter and a Palm Pilot. The custom built transmitters are placed strategically throughout a building. They feature a transmitting range of about 30 meters. Due to the design of the system, the information transmitted can be adapted according to the position (position accuracy is in the range of a few meters) and orientation of the user. A browser,

Figure 3.10: Palm Pilot and infrared transmitter in the IRREAL project (source Baus et al., visited 2005)

running on the PDAs, was developed to display both texts and graphics. The browser also features hyperlinks to allow the user to follow links between different topics. It is also possible to save data, to allow the user to review the information transmitted.

3.2.4.2 ARREAL

The ARREAL component (see Kruppa, 2001) realizes the outdoor navigation within the REAL resource-adaptive navigation system. The ARREAL prototype is realized using "off-the-shelf technology". It combines a small notebook, a standard GPS receiver, an electronic compass and a clip-on display. The notebook and the GPS receiver are integrated in a backpack (see Figure 3.11).

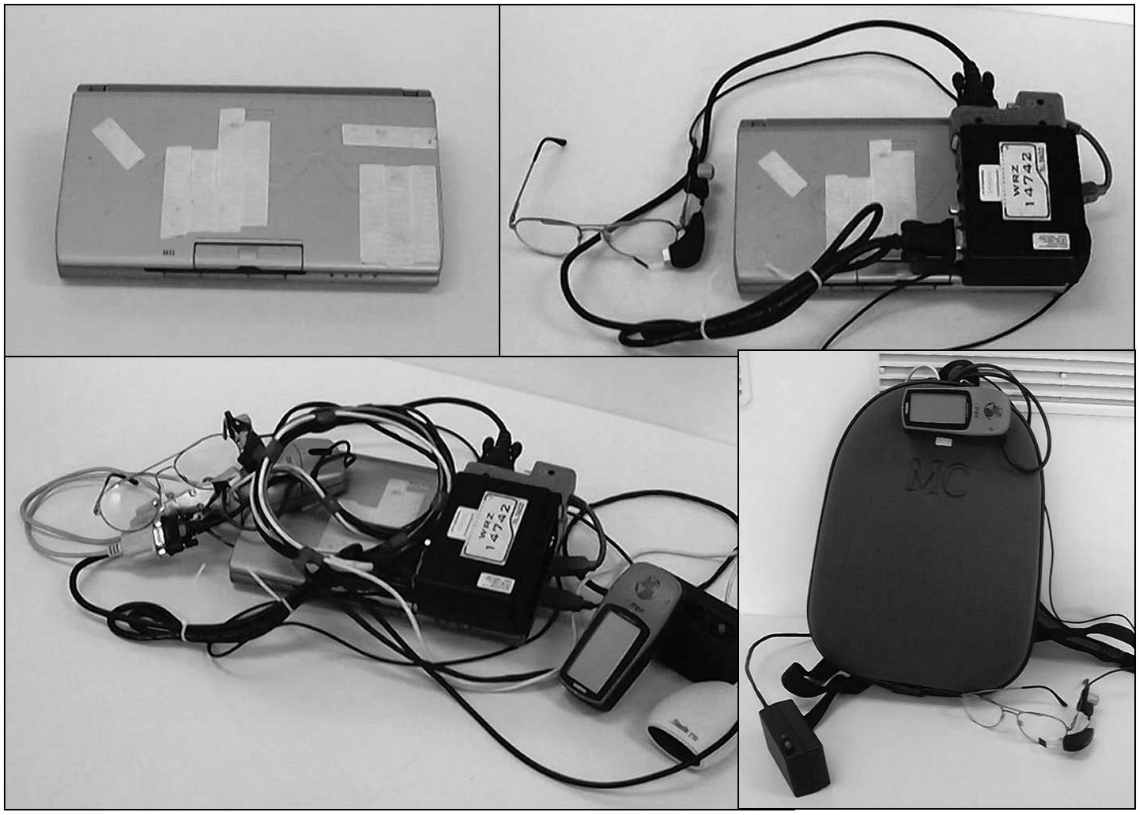


Figure 3.11: The hardware components of the ARREAL prototype (source: Kruppa, 2001)

The clip-on display features a small display which may be fixed on a arbitrary pair of glasses. The main benefit of such a display as compared to a PDA or laptop display is the fact that it allows a user to perceive information on the display while also monitoring the physical world. In addition, a clip-on display allows for hands-free operation which is very helpful in a pedestrian navigation system. The electronic compass is used to determine the orientation of the user. In addition, since the electronic compass is held in one hand by the user, it may also be used as a 3D pointing device, allowing the user to request information on particular physical objects by pointing at the object and pressing a button on the electronic compass.

The pedestrian navigation system supports two different modes:

- The navigation mode, allowing a user to state a destination and then subsequently guiding the user towards his/her destination by giving aural and visual cues

- The exploration mode, where users are free to move in any direction without stating a destination beforehand. In this mode, the system assists the user by offering information regarding objects which are physically close to the user (e.g. buildings and institutions located inside these buildings)

The information presented on the clip-on display is adapted depending on the cognitive load of the user and the different modes available. In exploration mode, the user is shown a detailed birds eye view with annotations on objects close to the user. During a navigation task, object annotations are reduced to those objects which are relevant to the task. The user is guided by verbal descriptions and arrows shown on the map, pointing towards the next waypoint (as shown in Figure 3.12). If the user is in a hurry (determined by the users actual speed based on the continuous GPS signal), the map will zoom out which results in a less detailed, but broader view.

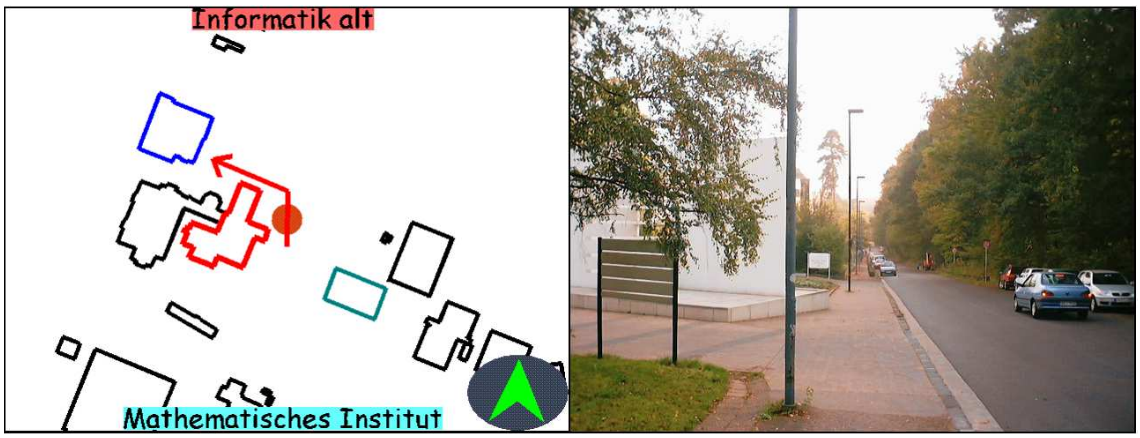


Figure 3.12: Exemplary screenshot of the ARREAL SYSTEM and the corresponding physical situation (source: Kruppa, 2001)

3.2.5 HIPS

The HIPS (**H**yper-**I**nteraction within **P**hysical **S**pace) project deals with the development of handheld electronic tour guides, allowing a tourist to navigate through physical space while receiving pertinent information. Application examples are city guides and museum tour-guides. The system provides contextual and personalized information based on the position detected by the system (see Broadbent & Marti, 1997). The user is guided by audio messages generated by the system (messages include instructions to find an item of interest, descriptions of items with references to items seen earlier and to others still to be seen). The presented information in HIPS is generated dynamically allowing to integrate user requests and preferences. The system is able to identify a situation and to select an appropriate information presentation according to the user's context. The system is based on a client-server structure. The clients (called *hippies*) are pen controlled palmtop computers, featuring a small screen and headphones. The clients are carried around by the user. To localize the clients, different techniques are used, namely GPS, infrared or radio frequency beacons. The client-server communication is realized via wireless communication technologies.

3.2.6 DeepMap

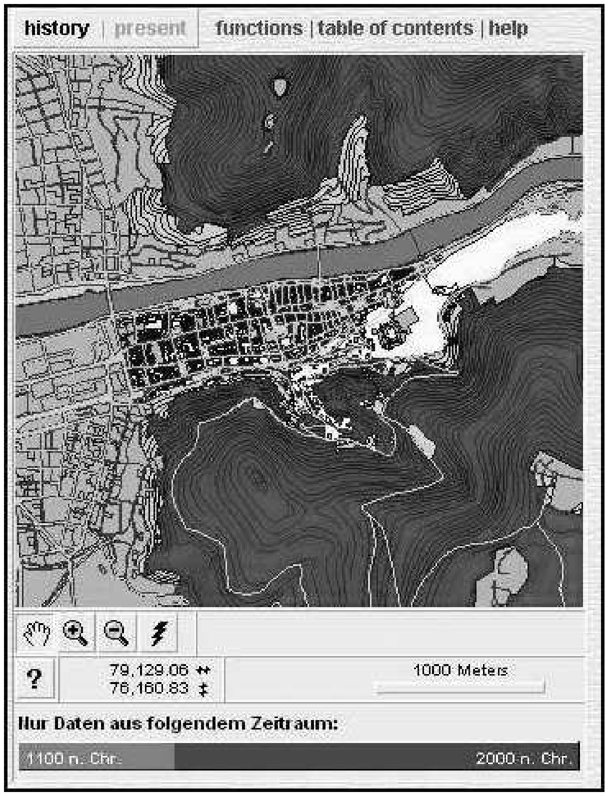


Figure 3.13: A screenshot of the DeepMap tourist guide (source: Malaka, visited 2005)

In the DeepMap project a digital mobile tourist guide is developed. The project combines several computer science research areas (Klaus Tschira Foundation, visited 2006):

- geo-information systems
- data bases
- natural language processing
- intelligent user interfaces
- knowledge representation

The project is divided into several sub-projects covering particular aspects of the final system. These sub-projects develop prototypes which are combined into the DeepMap system. The DeepMap - GIS (**G**eo **I**nformation **S**ystems) sub-project provides access to several different databases, like historical-, geographical-, and tourist-information databases. A virtual tourist guide (see Figure 3.13), in combination with a city guide for the Internet, enables the user to prepare his visit and plan a tour.

In the sub-project Integrated Map, methods to include external data sources into the DeepMap system are developed. These external data sources include information about hotels, restaurants, and cultural events. The information comes in many different formats. The goal is to enable DeepMap to make use of such information without depending on a certain database format. Another aspect of Integrated Map is to enable the user to use reservation services (e.g. ticket reservation or hotel reservation) via DeepMap.

The Talking Map sub-project deals with language processing for tourist information systems. Talking Map combines both recognition and generation of natural language. The recognition part is subdivided into a speech recognizer and a language understanding module, allowing the user to retrieve a desired information or to activate a certain task in a natural way. For example, the user can ask the system to navigate to a certain point of interest. The speech recognition is speaker independent. Talking Map features a spatial cognition engine, allowing the system to translate the spatial information contained in natural language utterances into a geometrical representation. The natural language generation allows the system to generate sentences including spatial information (e.g. “turn left and follow the street” . . .) as well as tourist-information.

The Virtual Map sub-project deals with the 3-dimensional reconstruction of Heidelberg (the city, where the DeepMap prototype is being tested and evaluated). The reconstruction includes both historical and actual states of the city. The Virtual Map sub-project uses computer vision technologies to support user localization and navigation tasks.

3.2.7 Sotto Voce

The basic idea behind the Sotto Voce project (see Aoki et al., 2002) is to build a guide system for a historical house taking into account the special needs of groups. Instead of solely providing information to individual users, the system is capable of supporting as well as encouraging communication among group members. In most cases, participants of docent-led tours are turned into a passive audience and audio tours often force users into isolation (Hood, 1983). In order to engage visitors into conversation, the electronic guidebook, forming the central part of the system, supports technologically mediated sharing of informational audio content. This sharing mechanism is called *eavesdropping* and will be explained later.

The Sotto Voce prototype consists of several guidebook devices. These devices combine a Compaq iPAQ 3650 hand-held computer featuring a high resolution color touch screen with a wireless local-area network (WLAN) card. In addition, headsets that do not fully occlude the ears and hence still allow users to communicate verbally with each other, are connected to the iPAQ. In order to support the audio sharing mechanism, two guidebook devices may be paired over the WLAN using standard Internet protocols (UDP/IP). Since the audio content to be presented is fixed and pre-installed on each guidebook, only synchronization messages need to be exchanged between the paired guidebooks (e.g. “start playing clip x”, “stop playing clip y”).

The user interface of the prototype is subdivided into two regions (see Figure 3.14). While a small region in the upper part of the touch screen features two different groups of radio buttons, one for controlling the eavesdrop mechanism which we will describe below and the other one which allows users to select the room in which they are located. The major, lower part of the touch screen is used to display a photograph of a wall of the room in which the user is standing.

Figure 3.14: The Sotto Voce prototype setup (source: Aoki et al., 2002)

By pressing hardware buttons on the iPAQ, users may change the wall displayed on the touch screen. The photographs resemble a visual interface with buttons to select audio content related to specific objects on the walls. The photographs displayed are a set of Hypertext Markup Language (HTML) image maps. If a user taps on the screen and hits the target region of an image map, the audio content corresponding to the tapped object will be started. If a user taps on a region which is not linked with an object, the guidebook displays transient target outlines appearing around the objects that may be tapped.

The eavesdropping mechanism allows two users of the system to share audio contents. When a visitor selects an audio clip by tapping on the touch screen, she always hears that particular audio clip. However, if the second person is playing an audio clip, while the first person is not currently listening to an audio clip, both users will hear the same audio clip. Audio clips are never mixed, and personally selected clips always take precedence. The playback of audio clips is synchronized, which means that, in the case where both users are listening to different audio clips, and one clip is shorter than the other, the corresponding user will not hear the other users audio clip from the beginning, but from the actual playback position. Users may control the eavesdrop mechanism by either turning it off or by selecting two different volume settings.

3.2.8 Summary

A central characteristic of a Migrating Character is its capability to guide a user in physical environments. It is hence necessary to supply a Migrating Character with positional information so that it may apply its navigational skills. Furthermore, the positional information is also necessary in order to allow the character to perform references to physical objects and more generally speaking, to allow the Migrating Character to provide location-based services.

As already discussed in Section 2.3, one of the main benefits of mobile computing technology, apart from the basic fact that it is available in a mobile context, is the possibility of offering location-based services. These services which are in one way or another directly linked to the users' (i.e. the systems') actual locations represent a new class of services which are driven through indirect interactions of the user with the system (e.g. the user's movement through physical space, interactions with physical objects). While in mobile navigation systems, the positional information is directly processed and integrated into the route planning and navigation task, it is also possible to link electronic information to physical location, as is the case in mobile tourist guides or also museum guides. By establishing this logical link between the virtual world and the physical world, new interaction metaphors for navigating through the virtual information space arise. Instead of offering all available information at every location, and, hence, possibly overstrain the user when trying to find a specific peace of information on a mobile device, the electronic information can be structured according to the physical layout of the environment. In this way, a pre-selection of relevant content which is directly linked to a specific physical location may be achieved. For example, in a cultural heritage setting (as in the case of the Sotto Voce project, see Section 3.2.7), information regarding certain exhibits is only relevant when standing close to the particular object, especially if the information to be presented directly refers to certain details of the exhibit which can only be explored, when standing right in front of it.

Depending on the different types of applications, the quality of the positional information needed for a specific service in order to work properly may vary. This is, for example, quite obvious when comparing a car navigation system with a pedestrian navigation system. While the car navigation may restrict the user's movements to streets within the systems maps (since cars are expected to only move on streets) regardless of the actual position derived from the positioning service, a pedestrian may basically move everywhere. Furthermore, on a pedestrian level, a few meters of difference in positional data may make a big difference (for example being on one side of the street or the other), while an positional inaccuracy of a few meters does not usually pose problems to a car navigation system. But even within the group of mobile applications especially tailored towards use by pedestrians or visitors of museums and exhibitions, the necessary quality of positional information may vary. When linking electronic information to physical locations it is important to be able to clearly distinguish between different locations being linked to different data. For example, in a museum, depending on the distance of one exhibit from the other, a high quality positioning system may be needed or sometimes a rough approximation may be sufficient. It is, however, important to either guarantee that a user's position will be available at any time while using the mobile system (which is almost impossible) or to provide some kind of backup mechanism which allows the mobile system to continue to work even though the functionality will probably be limited in such a situation.

3.3 Life-Like Characters Guiding Users in Virtual Reality Applications

The Migrating Character technology described in this work will allow life-like characters to accompany users while exploring physical environments. Furthermore, the Migrating Characters will be capable of guiding users through physical environments and they will especially be able to refer to physical objects by imitating the way humans perform such references (e.g. moving closer to the object, pointing at it, referring to it verbally, etc.). In this section we will review a number of projects which have realized such capabilities within life-like characters inhabiting virtual environments. However, the Migrating Characters are the first life-like characters transferring these features from the virtual to the physical world.

3.3.1 PPP Persona

The PPP (the Personalized Plan-based Presenter, see André, Rist, & Müller, 1997; André, Rist, & Müller, 1998a) system was developed by the German Research Center for Artificial Intelligence (DFKI). The main goal behind the project was to develop a personalized presentation agent, capable of providing multimodal instructions for the operation of technical devices. The PPP Persona combines spoken utterances (or written text in bubbles) with pointing gestures on visual objects on the screen in order to highlight certain details relevant to the current explanations given by the presentation agent. In order to highlight details which are out of reach of the presentation agent, the agent utilizes a virtual pointing device (i.e. a line drawn between the agents hand and the target object, see left-hand part of Figure 3.15).

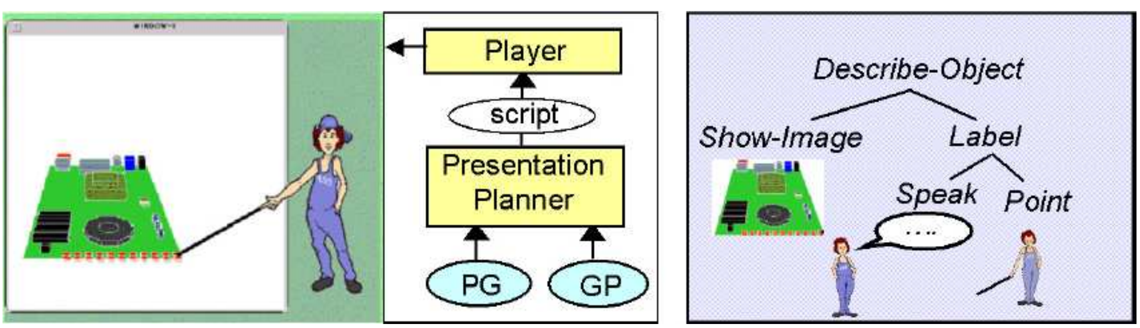


Figure 3.15: PPP Persona: screenshot and architecture (left), simple presentation plan (right) (source: Rist et al., 2004)

The presentations performed by the PPP Persona are script driven. These presentation scripts contain a specification of a multimedia presentation consisting of instructions for the presentation agent (i.e. text to be spoken or displayed, timing of gestures and utterances) and a temporal description of the sequence of multi-media objects to be displayed during the presentation. These scripts are automatically generated prior to the presentation execution. The generation process takes into account the presentation goals as well as generation parameters, such as time constraints regarding the presentation or the user's personal expertise in the current area of interest. The presentation planner starts the generation of the presentation script by decomposing the complex presentation recursively until all sub goals have been reduced to elementary production, retrieval, or presentation tasks. The resulting structure of this process is a hierarchical presentation plan (the right-hand part of Figure 3.15 shows a portion of a hierarchical presentation plan). This presentation plan is then transformed into a presentation script and forwarded to a player engine. This dedicated player engine is responsible for the synchronized playback of the agents animations and verbal speech output as well as the display of all involved multi-media objects and text elements. A symbolic representation of the script generation and playback process is shown in the middle of figure 3.15.

The PPP Persona was adapted to work in many different scenarios and applications. For example, the presentation agent was used to advertise accommodation offers from those found on the Internet in the WebPersona project (as described in André, Rist, & Müller, 1998b). After the user has stated an initial request, the system looks up a matching offer on the Internet. In addition, a presentation script is generated which later transcribes into a presentation performed by the PPP Persona, which uses gestures and speech to highlight specific details.

3.3.2 Cosmo

The animated pedagogical agent "Cosmo" (see Towns et al., 1998; Johnson, Rickel, & Lester, 2000) has been developed with a specific goal in mind: To improve deictic believability in animated agents inhabiting virtual environments. The idea is to allow an animated agent to refer to virtual objects in the same way humans would do it, by combinations of speech, locomotion and gestures. Implementing such human behavior in an animated agent living in a virtual 3D world should not only improve the believability of the agent, but it would also allow the character to disambiguate references to objects within the virtual world. Instead of relying on verbal descriptions alone, Cosmo is capable of moving towards the target object. In addition, it may also gaze into the direction of the object and it may also perform pointing gestures to further clarify the reference.

In order for the agent to be able to create utterances, deictic gestures and motions, it must have sufficient knowledge of the positions of objects within the virtual world. Furthermore, it must also know its own relative position with respect to the relevant objects and it must keep track of its previous explanations. To fulfill these requirements, Cosmo features a spatial deixis framework for achieving deictic believability. Within this spatial deixis framework, the planning mechanism for deictic gestures, locomotion and speech is subdivided into three phases.

In the first step, the ambiguity appraisal, the system determines whether a reference to the target object may be ambiguous. Based on an explanation plan which holds a record of objects the character has referred to in previous utterances within the same explanation sequence, the system calculates an initial potential for ambiguity when referring to the target object (i.e. the system determines whether the object was referred to within the last two utterances and whether there were other objects also referred to within these utterances).

The next phase is the gesture and locomotion planning. Based on the previously achieved ambiguity appraisal, the system employs a world model representing the relative positions of the objects in the scene to plan the character's movement and pointing gestures. The system decides whether it is necessary for the character to point at the target object. If so, the system then decides whether the character should move closer to the object, by taking into account the proximity of objects in the world. If the target object and another object are close to each other and if both were referred to in the last two utterances, the character will have to move closer to the target object. Each object within the virtual world is associated with an ontological category. If there are objects of the same category close to the target object, the character will also move closer to the target object. In addition, the character will also approach the target object, if the object is unusually small (this attribute is annotated in the world model).

The final step is the utterance planning phase. To produce a referring expression consisting of the appropriate proximal ("this" or "these") or non-proximal ("that" or "those") demonstratives and pronouns, the deictic system considers the world model, the ambiguity assessment and focus information. The resulting utterance is also influenced by the relative locations of the target object and the virtual character.

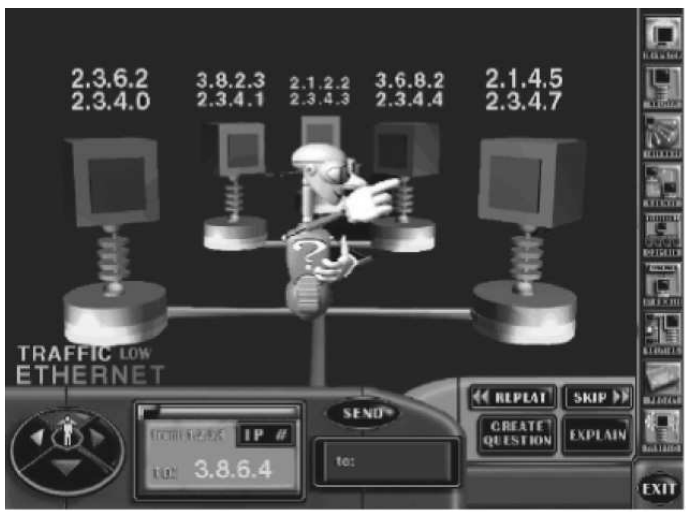


Figure 3.16: Cosmo and the INTERNET ADVISOR world (source: Lester et al., 1999)

As a test bed for the Cosmo character, a learning environment was chosen. In order to foster the evaluation of the deictic believability of the character, the learning environment should be populated by many similar objects and it should involve a problem-solving task requiring students to make decisions based on factors included in the environment. The Internet Advisor application (see Figure 3.16) offers the required parameters. The virtual world in the Internet Advisor consists of many routers and networks. Integrated in this virtual world is the Cosmo character which interacts with the students while they learn about network routing mechanisms. The student's task, after an initial explanation by Cosmo, is to escort a packet through the network by navigating through a number of sub-nets. During the initial explanation and whenever the student is in need of advice (which is determined by analyzing the decisions the student takes), Cosmo provides hints and explains fundamental concepts by highlighting particular routers or networks. Within this setting, disambiguation of object references is very important.

3.3.3 WhizLow

The explanatory life-like avatar WhizLow (see Lester, Zettlemoyer, Grégoire, & Bares, 1999) inhabits a 3D learning environment called CPU CITY (see figure 3.17). The main purpose of this 3D learning environment is to teach students (non-technical novices) the fundamental principles of computer architecture and systems. The CPU CITY 3D world is a simplified 3D model of a motherboard consisting of three main components: the random access memory (RAM), the central processing unit (CPU) and a hard drive. These different components are represented as simple 3D geometric objects within the 3D world and they are connected with each other via buses (represented as streets in the 3D world). The resulting virtual world resembles a cityscape. The purpose of the life-like avatar WhizLow is to perform actions within the 3D world which are

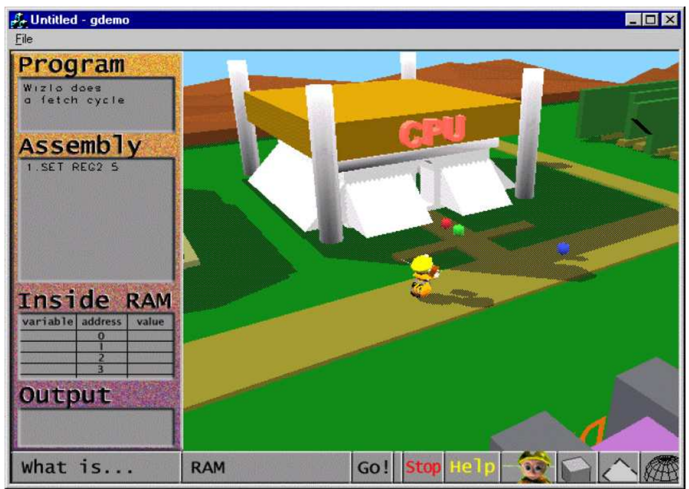


Figure 3.17: WhizLow in the CPU CITY 3D learning environment (source: Johnson et al., 2000)

specified by the students in a high-level programming language. These high-level task descriptions are then, in a number of consecutive steps which we will describe in detail later on, transferred into a usually several minute demonstration in which WhizLow performs all the necessary steps to fulfill the task specified by the student. In addition, the avatar also fills in helpful explanations regarding the actions he performs in order to achieve the goal stated by the student.

The transformation of the initial task description given by the student to the final demonstration performed by WhizLow is initialized by a transformation from high-level task description to a task tree whose leaves are action specifications for the character to perform. In order to do so, a goal decomposition planner takes into account a rich representation of task knowledge. The action specifications at this stage are however only recommendations for the avatar which still need to be interpreted within the physical realities of the 3D learning environment.

The next step is the explanatory task interpretation. Based on the sequence of action descriptions calculated in step one, the interpreter utilizes its 3D world knowledge in order to determine the navigation, manipulation and verbal behaviors for the WhizLow avatar. Navigation behaviors are based on the physical layout of objects in the virtual world and, hence, involve a route planning mechanism. Manipulation behaviors relate to single virtual objects which are to be manipulated during the planned task and verbal behaviors are sequenced to create the accompanying explanations.

Finally, a narrative cinematography planner is used to determine an optimal sequence of shots, zooms and pans for the virtual camera that “films” the avatar while performing the user specified task. The resulting presentation, as in contrast to directly controlled life-like characters (for example Steve, see Section 3.3.4), enables students to observe an life-like avatar while it takes all the detailed actions for the student. Thus, students can attend to the concepts in the domain.

3.3.4 STEVE

The animated agent STEVE (Soar Training Expert for Virtual Environments, see Rickel & Johnson, 1999 and Marsella, Gratch, & Rickel, 2004) was especially designed to support students in learning how to perform physical, procedural tasks. STEVE is capable of achieving this goal by either demonstrating how to perform a certain task or by monitoring the student while performing the task and by giving comments and advices according to the students decisions. Steve inhabits virtual 3D worlds featuring virtual objects, possibly other animated agents and human users. The users are experiencing the virtual 3D world by using a combination of a head-mounted-display, several position sensors on the users heads and hands, a data glove, a set of headphones, and a microphone. This hardware combination allows users to interact naturally with the virtual world by pointing at objects, grabbing objects and by using spoken language to communicate with STEVE. In order to allow STEVE to successfully work in task-oriented collaboration settings, the agent requires the capabilities of perceiving the state of the virtual world and assessing the state of goals. In addition, the character needs to be able to construct plans to achieve certain goals and then it should be capable of autonomously navigating through the virtual 3D world to execute its plans.

The architecture of STEVE is divided into three major components: a perception module, a cognition module, and a motor control module. The perception module monitors messages from a message dispatcher, which is used as a communication gateway between all components allowing each component to register for specific types of messages and to automatically identify those messages that are relevant for STEVE. These messages carry information about changes in the state of the virtual world which are the results of actions of either human users or animated agents. For example, by pressing a button, a light may be turned on or off. The perception module, after parsing these messages, forwards the relevant information to the cognition module. The job of the cognition module is to analyze these data and to choose appropriate goals accordingly. Based on these goals, the cognition module constructs and executes plans and sends commands to the motor control module. The motor control module controls the agent’s body by translating the higher level commands received from the cognition module into a sequence of animation primitives which are sent via the message dispatcher to the simulator for final execution.

As an application example, STEVE was integrated in a scenario where students have to learn how to operate and control machines aboard a ship (Figure 3.18 shows STEVE demonstrating how to inspect a high pressure air-compressor). Each lesson allows for different interactions and the flow of the whole session is open ended. At the beginning of each session, STEVE explains the task to be completed during the session. After this introductory part, STEVE starts to demonstrate all the necessary steps in order to achieve the actual goal. However, the students are always free

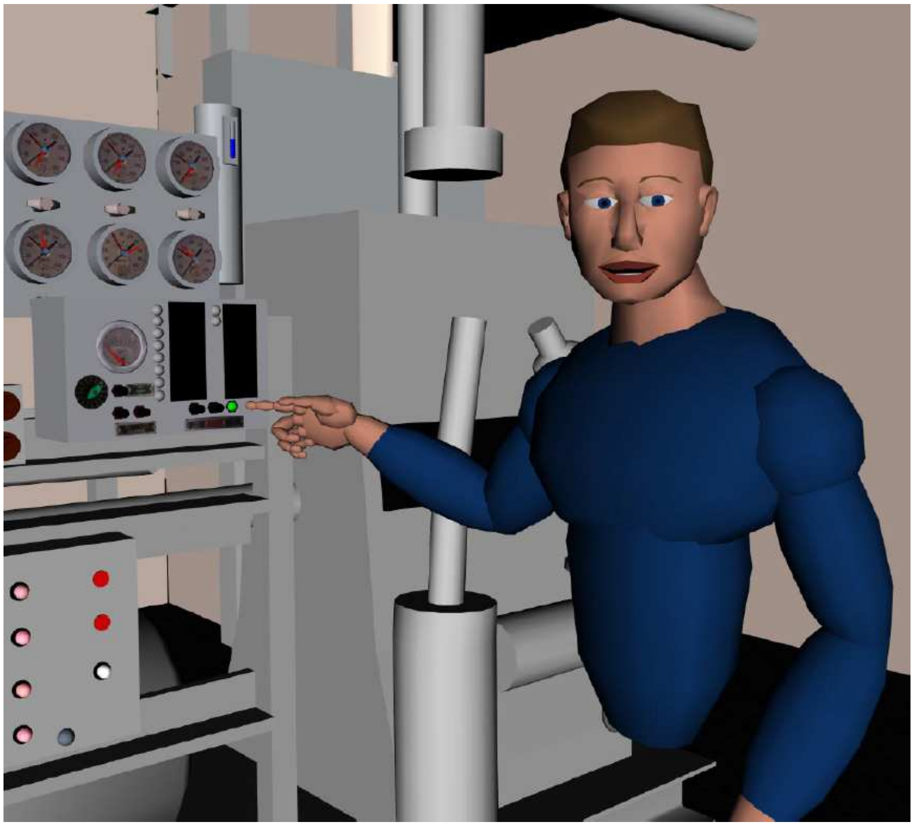


Figure 3.18: STEVE describing a power switch (source: Marsella et al., 2004)

to interrupt STEVE for a question (e.g. “Why do I have to press this button?”) or to finish the task themselves. In case a student decides to finish the task by himself, STEVE switches to a monitoring mode. In this mode, STEVE analyzes each action performed by the user and gives comments and offers advice if necessary.

3.3.5 Summary

Even though the projects described in this section are quite different from one another, they share a common goal, namely to allow a life-like character to guide a users attention towards specific virtual objects located within the virtual world the character inhabits. The various application scenarios as well as the different virtual worlds have lead to several approaches in order to realize the desired capabilities within a life-like character.

The two-dimensional world of the PPP Persona (see section 3.15) lead to a solution enabling the character to perform simple pointing gestures in combination with utterances shown in bubbles. If the target object is out of reach for a pointing gesture, the PPP Persona utilizes a simple, virtual pointing device in order to highlight a certain detail. The presentations performed by the PPP Persona are script driven and include a sequence of annotated media objects which is synchronized with the characters explanations. As opposed to the other three projects reviewed in this section, the PPP Persona inhabits a virtual world which does not require complex character movements. Instead of having characters moving through the virtual world, the virtual world is adapted in each presentation step (i.e. media objects appear and disappear close to the characters location on the screen).

All other life-like characters presented in this section inhabit three-dimensional worlds. When either of these characters needs to refer to a particular object within the virtual world, it first needs to ensure that it is close to the object. If that is not the case, it has to automatically determine a way to move close to the object first. Moving through these three-dimensional worlds does not only require a route finding algorithm but it also demands a complete knowledge covering all the objects as well as all the pathways within the 3D-world. Furthermore, the characters also have to decide, how close to move to an object and which type of reference to perform. The Cosmo character (reviewed in Section 3.3.2, for example, is capable of automatically deciding, whether a reference to a particular object within the 3D world demands one of the following actions:

- Relocation of the character to another virtual room or somewhere closer to the relevant object
- A pointing gesture towards the object
- A head movement and a gaze towards the referred object
- A spoken utterance identifying the object

A complex set of rules, in combination with a very detailed world knowledge database is employed by the character in order to plan its own behavior during presentations which include references to virtual objects. The rules which come to use resemble the way in which humans refer to physical objects. Humans take different strategies for object references, depending on the type of object they refer to as well as the context in which the reference occurs. Sometimes it may be sufficient to describe the object verbally, for example, when referring to a car by saying: "The green car". However, if this reference is possibly ambiguous, because there are other green cars around, different strategies may be applied to further clarify the reference. One solution is to expand the verbal description by adding further details, e.g.: "The green convertible". Another option would be to use a pointing gesture while saying, "The green car over there". If all these strategies fail, a last solution is always to move closer to the object in order to clarify a pointing gesture.

Also especially important is the capability of humans to adapt to a changing context. If an object (e.g. a small indicator light) can change its color, these changes have to be reflected when referring to the object. The animated agent STEVE (see Section 3.3.4) is capable of constantly monitoring the status of the virtual world it inhabits. If the status of some object is changed, or if an object is removed or added to the scene, these facts are integrated in a status model of the world. In the STEVE prototype, both STEVE himself and the users are allowed to use objects within the virtual world and thereby change the status of the world. These changes are constantly monitored by STEVE and reflected in its own behavior.

3.4 Life-Like Characters in Mixed Reality Applications

The life-like characters described in the last section all share a common peculiarity, namely the fact that they are limited to virtual reality scenarios. The Migrating Characters take a different approach which will allow them to work in virtual as well as in physical environments. However, there is not only the completely disjoint worlds of virtual and physical reality, there is also something in between. According to Milgram and Kishino (1994), there exists a virtuality continuum (as illustrated in figure 3.19). Milgram and Kishino define everything which is in between real environment and virtual environment as *mixed reality*.

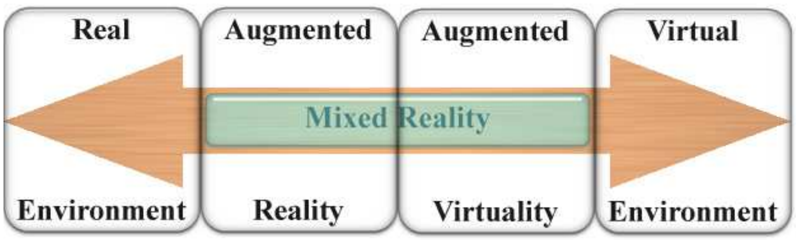


Figure 3.19: The Virtuality Continuum (according to: Milgram & Kishino, 1994)

Mixed reality is hence either augmenting the physical world with virtual objects or the other way round, augmenting virtual worlds by including physical objects (or representations thereof). André, Dorfmüller-Ulhaas, and Rehm (2005) transfer the virtual continuum by Milgram and Kishino to the research area of life-like characters. In this section we will review projects which aim at the realization of life-like characters in mixed reality applications. While in the first project, MACK, the mixed reality is constituted by a virtual character in a virtual world, performing 3D pointing gestures towards objects from the real world, the second project goes a step further. By using augmented reality technology , the Virtual Augsburg prototype allows a user to interact with objects in the physical environment while also interacting with a life-like character being overlaid on the real world image. Furthermore, the user may indirectly control the character by manipulating physical objects.

3.4.1 MACK

The MACK (Media lab Autonomous Conversational Kiosk, see also Stocky & Cassell, 2002) project, conducted at the Massachusetts Institute of Technology (MIT), aims at the realization of a embodied conversational agent capable of assisting a user in finding a particular location in a building by means of direction giving. The agent utilizes a combination of speech and gestures to perform and clarify the direction giving. For user input, the agent also realizes a multimodal approach, allowing users to interact with the system by means of pointing gestures combined with natural language. The communicative capabilities allow a user to interact with the agent without any need for training, since the utilized communication protocols are commonly known and are used by ordinary users in daily life.

The agent itself is a 3D animated blue robot located on a fixed stationary information kiosk. In order to achieve the illusion of maximum immersion of the agent in the physical environment, the kiosk features a camera oriented towards the back of the kiosk. The video stream taken by the camera appears behind the agent by using a video mixer which is also integrated into the hardware setup of the environment. When looking at the virtuality continuum of Milgram and Kishino (see Figure 3.19) we see that the approach used in the MACK prototype is located in the area of augmented virtuality. Figure 3.20 A shows the agent on the kiosk with the aforementioned immersion technology.

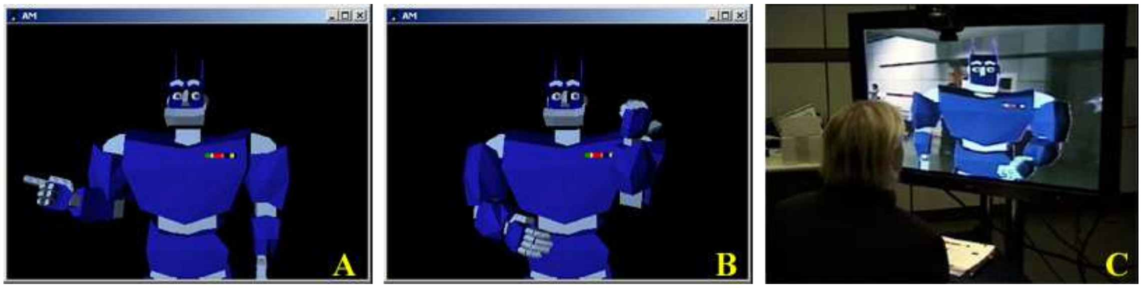


Figure 3.20: MACK performing pointing gestures and interacting with a user (source: Stocky, 2002)

In order to allow for the multimodal input by the user, MACK employs speech recognition in combination with two additional sensors. While the first one, a pressure-sensing chair mat, is used to detect user presence, the second one, being an embedded WACOM tablet (a touch sensitive device used with a stylus), is used to detect user gestures on map. The map is located on the WACOM tablet and a user may hence produce a multimodal interaction, for example, by saying: “How do I get here?” and then pointing with the stylus on the desired location on the map.

The reactions of the agent are performed in a similar fashion. The agent may utilize a combination of spoken utterances and pointing gestures. The gestures performed may either happen on the screen of the kiosk, by means of a 3D pointing gesture performed by the agent (see Figure 3.20 A and 3.20 B for two examples pointing gestures), or it may happen in the real world, by using an LCD (Liquid Crystal Display) projector mounted on top of the map to highlight a specific detail or region. The gestures and utterances are calculated according to the agent’s own location and orientation which are known to the system.

The agent progressively uses three different methods to disambiguate directions given:

1. Relative descriptions like: “Do you know the great statue at the main entrance?”, “It’s right there!”
2. Explanations including deictic gestures like: “Go through that door...“ while pointing at the mentioned object
3. Map-based route planning like: “Go to the elevator first...” while highlighting the object on the map

Stocky and Cassell (2002) base their decision of using these three methods for the MACK agent on the results of a user study during which participants were asked to find their way to two different locations by asking passersby. The location chosen for these interactions allowed the participants and passersby to utilize a map. The results of the study indicated that humans tend to use the three strategies mentioned above for direction giving, usually with a tendency to try relative descriptions first and only utilizing the remaining strategies if the first one fails. This typical behavior was then realized in the MACK prototype.

3.4.2 Virtual Augsburg

The Virtual Augsburg project (see André, Dorfmüller-Ulhaas, & Rist, 2004) conducted by the *Multimedia Concepts and their Applications group* at Augsburg University, is taking a very different approach towards the realization of a life-like character in a mixed reality scenario. Instead of augmenting the virtual environment of the character with representations of physical objects, as in the case of the previously mentioned MACK project (see Section 3.4.1), the synthetic character Ritchie, which was developed for the Virtual Augsburg project, is augmenting the physical environment of the user. So, in the sense of the virtuality continuum of Milgram and Kishino (see Figure 3.19), the Virtual Augsburg project (i.e. the synthetic character Ritchie) falls into the area of augmented reality. In a prototype application, a table-top application is jointly explored by the user and the synthetic character Ritchie. The application combines a real city map of the city of Augsburg, being laid out on the table, with overlaid virtual 3D buildings of the city center of Augsburg (see figures 3.21 and 3.22).

Figure 3.21: Virtual Augsburg: Physical interaction and tracking system (source: André et al., 2004)

In this way, the user and the character share a common physical space allowing for a high degree of immersion . The user gets the impression of the character perceiving the scene in the same way he/she does, resulting in a stronger feeling of a shared experience between the user and the synthetic character.

From a technical point of view, the approach presented by André, Dorfmüller-Ulhaas, and Rist demands high amounts of computational power. The virtual objects (including the synthetic character Ritchie) are overlaid on the user's field of view by means of a head mounted display (see the right-hand side of figure 3.21). Located on top of the head mounted display is a camera fulfilling two different purposes. Firstly, the video stream captured from the camera is shown in the head mounted display, allowing the user to perceive the physical environment while wearing the non-see-through head mounted display. Secondly, the camera's video stream is analyzed in order to detect visual markers on the map on the table. The real-time detection of these markers is necessary, in order to calculate the locations where the virtual 3D objects should be placed. The technology allows to place virtual objects in a scene which complements the camera image shown in the head mounted display in a consistent, visually correct way. Furthermore, detecting the markers allows the system to calculate the user's attentional focus which consequently influences the behavior of the synthetic character. The character tries for example to detect, whether the user is following its movements and undertakes necessary counter measures if the system detects that the user is not following the character.

The user may interact with the synthetic character by manipulating a physical object. The object, being a frustum of a pyramid, has a visual marker on each side. Each marker triggers a different action of the user. It is, for example, possible to place the frustum somewhere on the table, and once the marker is detected by the system, the character will calculate the shortest path from its actual position to the position of the detected marker and it will start walking towards it (see also Dorfmüller-Ulhaas & Andre, 2005).

The synthetic character is however not fixed to the augmented reality scenario described above. Instead, the same character may also be used in a virtual reality environment (see left-hand side of Figure 3.22). In the virtual environment, as in the augmented reality scenario, the character may freely move within the scene. The idea behind this additional scenario is that of providing a traversable interface (Koleva, Schnädelbach, Benford, & Greenhalgh, 2000). Traversable interfaces address the problem of allowing humans as well as life-like characters to move across the border in between virtual and physical world. This will allow both users and characters, in case the information presentation requires so, to move from interaction space to another. In the case of the synthetic character Ritchie, the traversable interface allows a user to first follow Ritchie on a tour through virtual Augsburg and in a second step, to explore the miniaturized model of the city laid out on a table (as described above).

Figure 3.22: Virtual Augsburg: A character in virtual and mixed reality environment (source: André et al., 2004)

3.4.3 Summary

The projects reviewed in this section take two very different approaches towards the realization of a life-like character in a mixed-reality scenario. While the MACK project (see Section 3.4.1) realizes a mixed reality scenario by integrating parts of the physical world into the virtual world in which the character resides, the Virtual Augsburg project (see Section 3.4.2) integrates objects from a virtual world (including a three-dimensional life-like character) into the perceived physical world around the user. The first approach is hence augmenting virtuality while the second approach is augmenting reality. One goal, which is shared among both projects is that of increasing the user's immersion into the mixed-reality world created. Apart from the level of user immersion, the chosen approach also has further implications for the realized systems. While an augmented virtuality approach, even though the character may seem to become part of the physical surroundings, does still fix the character to a device and consequently to the devices' location, an augmented reality approach actually allows a life-like character to move freely through three-dimensional space within the borders posed by the technical setup. The necessary technical setup for an augmented reality scenario on the other hand may limit the usability of the character. This is due to the fact that a user needs to be equipped with quite a number of appliances which usually feature cable which are connected to some stationary device. Hence, the character gains the freedom to move in three-dimensional space, while the users' freedom to move is restricted by the technical setup. Also, as in the setup of the Virtual Augsburg project, the augmented reality is actually generated by taking away the physical reality perception complete from the user and instead, integrating the missing visual information about the physical surroundings by means of a constant camera stream integrated into the virtual world displayed in the head mounted display. Since the hardware available as of today (meaning both cameras and head mounted displays) may not produce a visual representation of the physical surround matching the capacities of the human visual perception system, the resulting mixed-reality integrates only a "filtered" physical world.

The Migrating Character technology takes yet another approach towards the realization of life-like characters in mixed-reality. Instead of either augmenting reality or augmenting virtuality, the Migrating Characters may actually do both. Firstly, a Migrating Character on a mobile device may use the augmented virtuality approach, by, for example, integrating visual representations of real physical objects into the character presentations (see also Sections 4.2.1 and 6.1). The immersion of the character with the physical world is further improved by the contextual knowledge derived from additional sensors which consequently changes the characters behavior (which we will discuss in detail in Section 4.1.3). Secondly, a Migrating Character may become part of the physical environment by realizing it as a projection on a wall. Furthermore, using a steerable projector (an approach we will discuss in Section 6.2), the character may be enabled to move freely among arbitrary walls within the physical environment. In this case, as opposed to the augmented reality approach in the Virtual Augsburg project, the character is simply added to the physical world without taking away the "real" visual perception of the surroundings from the user. Since the use of the projection technique does not require a user to wear any hardware itself, this approach also will not limit the freedom of locomotion of the user.

3.5 Combined use of Mobile and/or Stationary Devices

The Migrating Characters described in this work are not limited to either mobile or stationary devices. Instead they are capable of using both classes of devices and furthermore they are capable of using combinations of different devices. In this section we will review projects which have realized applications that involve multiple input and output devices. These device combinations are used for many different purposes. For example, when combining a mobile and a stationary devices, the mobile device is often used as some kind of remote control for the stationary system. In the PEBBLES project (see section 3.5.1), several different ways of remotely controlling applications by means of a PDA are explored. In the Situated Computing Framework (described in Section 3.5.2), a hand-held device is used in order to access multimedia contents and services which are consequently shown on a composite device in the vicinity of the user. Another way of using combinations of input and output devices is realized by logically combining the displays of the different devices. In this way a display continuum is generated. The display continuum is not simply a larger display device, but often also offers advanced functionalities. In the Communication Chairs project (see Section 3.5.3), users are free to move visual representations of objects from one screen to another. This way, the data which is linked to the visual representation is transferred from one device to another.

All of the projects mentioned above share a common peculiarity. They all use device combinations in a parallel way as opposed to using them in a sequential way. However, there are also many approaches which, instead, support a sequential use of different devices and device classes. For example, both the HIPPIE project (reviewed in Section 3.5.4) and the REAL project (see Sections 3.2.4 and 3.5.5) allow a user to prepare a future task by using a stationary device and later on, a mobile device is used to actually fulfill the task. In the case of HIPPIE, the user prepares a museum visit by pre-selecting content of special interest and the system thereby generates a personalized museum tour on a mobile device to be used during the actual museum visit. In REAL, the user prepares a navigation task by selecting a destination and watching the automatically generated presentations illustrating the way ahead. Once finished, the user is guided to the desired destination by means of a PDA.

3.5.1 PEBBLES

In the PEBBLES (PDAs for Entry of Both Bytes and Locations from External Sources, see Myers, B., 2001) project at Carnegie Mellon University, different methods for spreading computing functions as well as the related user interfaces across all computing and input/output devices available to the user are being examined. The main focus of the project is to find out how handhelds and PCs may work together when both are available. During the course of the project, a number of applications have been developed allowing for a combined use of a handheld and a PC. The handheld is connected to the PC using either a cable, an infrared connection or a wireless connection (Bluetooth or Wireless LAN) and is being used for various purposes. Some of these are:

- Slide Show Commander, allowing the user to control a Microsoft Powerpoint Presentation
- Remote Commander, using the handheld as a keyboard and to control the mouse

- ButtonScroll, using the handheld to control scrolling on the PC
- Scribble, allowing several users to directly "scribble" on the display of a single PC by utilizing a unique cursor for each user, controlled via the PDA of the corresponding user
- ShortCutter, using the handheld to draw interfaces which control remote applications on a PC

The PEBBLES project is investigating the many ways that handhelds can serve as a useful adjunct to stationary computers. In PEBBLES, two or more devices are used at a time to improve the usability of the combined system. The display of the PC is used to present information, while the display of the handheld is used for user input purposes.

3.5.2 Situated Computing Framework

At Siemens Corporate Research a Situated Computing Framework (SCF, see Pham, T., Schneider, G., & Goose, S., 2000) has been developed. The goal of this framework is to implement methods that allow users a high degree of mobility without restricting the access to rich multimedia content due to the screen size of most mobile devices. A distributed computing concept offers classes of ubiquitous and mobile multimedia services to small screen devices. The SCF uses the handheld as a unique access device allowing users to request multimedia information and services. The SCF handles such a request by determining the desired multimedia content and presenting it on the most appropriate composite device in the vicinity of the user. The SCF is aware of the multimedia capabilities of each device and hence may extract appropriate data from its structured databases. The handhelds are not only used to handle user input, but they also detect composite devices nearby and report this information to a central server. The SCF combines the handheld with several composite devices located in the environment, featuring different multimedia capabilities. Two devices are used at a time. While the composite device is used to playback multimedia content, the mobile device is used to control the playback.

3.5.3 CommChair

In the CommChair project (see Müller-Tomfelde C. & Reischl W., 1998) an office chair was developed, featuring either a pen-based computer display or a laptop docking facility integrated into the chairs armrest. The room in which the chairs reside, is equipped with sensors allowing the system to determine the position of each chair as well as the persons sitting on the chairs. In this way, a personalized environment may be presented to each user. The chairs are connected to the Ethernet network via wireless LAN. Shared information displays may be set up by simply moving two or more chairs together. This way, users may share information and data located on their chair by simply grabbing it on their display and moving it to another chairs display. In an example scenario, the CommChairs were used in a meeting room to control a commercial client-server software for brainstorming which was run on a remote interactive electronic wall. The CommChairs combine the idea of using a small display as a remote control of a large display and the idea to use two or more displays as a combined/extended display.

3.5.4 HIPPIE

The HIPPIE system is an Internet-based museum guide which may be used at home as well as at the physical museum (see Oppermann & Specht, 2000). While in the museum, a hand-held computer (Toshiba Libretto 100 CT, including wireless LAN connectivity) is used. The information presentation on the mobile device takes into account the actual user location as well as the users knowledge and presentation preferences.

The basic idea behind the HIPPIE museum guide is a process-oriented approach. Instead of concentrating solely on the actual visit in the museum, in HIPPIE the visit starts with a preparation phase at home, which is followed by the physical visit. After the visit, an evaluation phase at home concludes the experience. All this is possible due to the nomadic characteristic of the approach (discussed in Oppermann, Specht, & Jaceniak, 1999), which allows for access to the user's personal information space at arbitrary locations and independent of specific hardware.

Before visiting a museum, the HIPPIE system offers users the opportunity to prepare their visit individually at home via the Internet. Users can browse through the contents of the museum, they may prepare personal tours and they may also mark object they are especially interested in. All the information gathered during this preparatory process is stored online and made available for later use by the mobile system in the museum.

During the actual museum visit, the mostly visually oriented content of the physical environment is complemented by an augmented reality components and audio output. Based on the user's location, determined by a combined system of infrared beacons sending a unique ID and an electronic compass, the system identifies relevant objects close to the user and chooses multimedia content to be presented accordingly. In addition, the information selection is also influenced by a user model including a representation of the user's knowledge as well as the user's personal presentation preferences. This user model is initialized during the preparation phase at home and is constantly updated by the system during the visit by monitoring and interpreting the user's interactions with the system as well as the path the user chooses within the museum.

The information gathered by the system during the user's visit in the museum is again stored online. This information is available for the user during the evaluation phase at home. This additional service allows the users to call to mind perceived information as well as allowing them to search for additional information. Users also have full access to the complete information space and to seen objects.

3.5.5 REAL (Revisited)

As described in Section 3.2.4, the REAL pedestrian navigation system utilizes a PDA to guide users in both indoor and outdoor navigation scenarios. However, the system also allows users to prepare the actual navigation task by using a large stationary information system. The stationary system, featuring a large display as well as advanced multimedia capabilities, allows users to state their destination. Based on this information, the stationary system plans a path starting at the actual position. Due to the fact that the mere textual descriptions of spatial information have proven to be inadequate quite frequently (for example, discussed by Towns, Callaway, & Lester, 1998), the stationary system offers a whole repository of visualizations illustrating the planned path (for a detailed description of all the available visualization techniques see Baus, 2003). Among these different visualizations are both 2D and 3D animations as well as still images (figure 3.23 shows still images of a 3D animation including a 3D avatar playing the role of the user navigating through the building).

Figure 3.23: Path visualization utilizing a virtual character as an avatar (source: Baus, 2003)

The idea behind these path visualizations is to support users in remembering the way-description easier. Included in these visualizations are hints to landmarks on the way to the users destination. Recognizing these landmarks later on during the real navigation task will help users to remember the planned path even better.

Snapshots of the path visualizations may also be shown on the mobile device again during the navigation task. Due to the different capabilities of the mobile and stationary hardware, the quality of the images is adapted accordingly. Figure 3.24 shows a photo of a real environment as well as two different graphical abstractions generated by the REAL system.

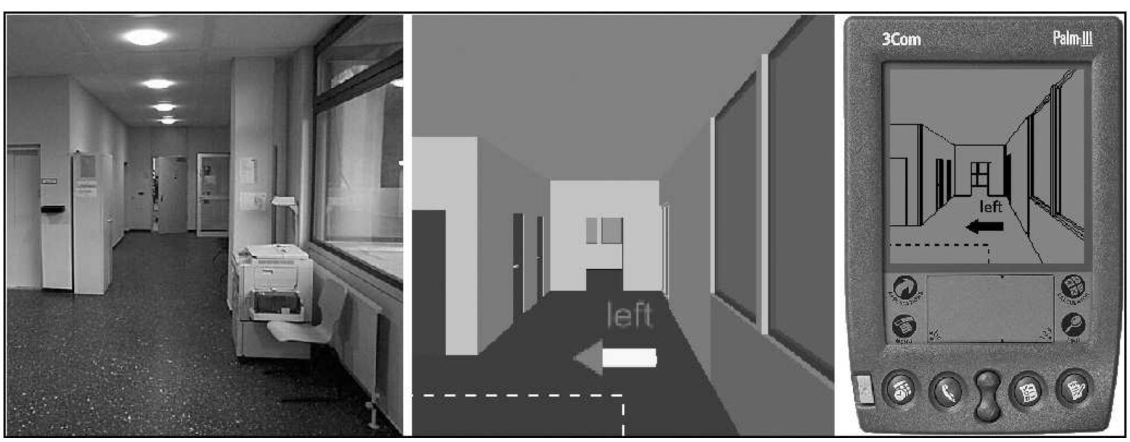


Figure 3.24: Real environment and generated graphical abstractions in REAL (source: Lohse, 2001)

3.5.6 SmartKom

In SmartKom (see Wahlster, Reithinger, & Blocher, 2001b; Wahlster, Reithinger, & Blocher, 2001a) an anthropomorphic and affective user interface combining speech, gesture and facial expressions for input and output is being developed. The anthropomorphic interface is realized by an interface agent called Smartakus. The interaction technology developed in Smartkom is based on the situated, delegation-oriented dialog paradigm . Instead of forcing the user to take each step towards the completion of a certain task by himself, this paradigm allows users to delegate high-level tasks to a virtual communication assistant.

The user interface in SmartKom allows for all possible combinations of speech and gestures in order to communicate with the system. A user may, for example, combine an utterance (defining the intention of the user) with a pointing gesture on a touch screen to define the object of the intended action (as illustrated in figure 3.25). In case the user input is ambiguous and more precise information is necessary, Smartakus is capable of requesting the additional input by means of a natural language dialog incorporating gestures and facial expressions. SmartKom supports multilingual input and output using speech synthesis and speech recognition modules for German and English.

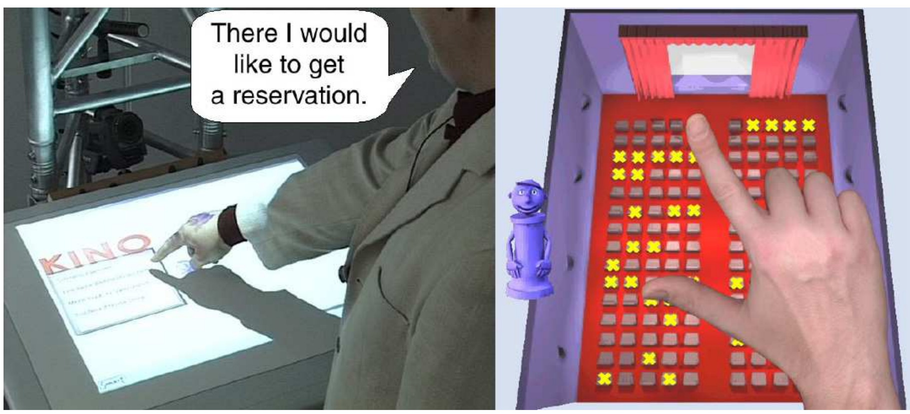


Figure 3.25: Multimodal interaction with the SmartKom system (source: Wahlster et al., 2001a)

The SmartKom system has been developed for three different classes of devices and corresponding application scenarios:

- SmartKom-Public is a multimodal communication booth combining an LCD projector, video-based gesture recognition hardware, a projection panel and a close-range microphone. The graphical user interface is projected by mmeans o a projector mounted above the projection panel. The system's primary purpose is to offer a platform to access information on specific topics. The implemented demonstrator offers access information relevant to tourists. For example, information on hotels and restaurants.
- SmartKom-Mobile is a mobile device version of the SmartKom system offering an extension to standard car navigation systems by allowing users to access the Internet in a convenient way. In addition, the mobile version of SmartKom offers route planning and an interactive

navigation through a city for pedestrians. The SmartKom-Mobile system will be reviewed in more detail in section 3.6.1.

- SmartKom-Home/Office offers a unified access to standard services like phone and e-mail as well as an easy way to control consumer electronics. For example, using the electronic program guide for television, the system allows users to program their videocassette recorder.

All the different SmartKom versions share a common architecture in the background. The architecture is organized in a modular way and combines up to 40 components. These components are organized into different subgroups related to specific tasks, namely input devices, media analysis, interaction management, application management and media design. Based on a multi-blackboard architecture with parallel processing threads, the different modules exchange data formatted in a new, XML-based markup language called M3L (MultiModal Markup Language). M3L was designed to support the representation and exchange of multimodal content, synchronization and confidence in processing results.

3.5.7 AgentSalon and PalmGuide

In the AgentSalon project (see Sumi & Mase, 2001b; Sumi & Mase, 2004), several mobile devices (running the PalmGuide system described in Section 3.6.2) may be used in combination with a single stationary system, a so-called information kiosk which is assumed to be located in a meeting place of an exhibition site. The main idea behind the system is to support face-to-face discussions and exchange of knowledge by tempting users to chat with each other. In order to achieve this goal, users are monitored while exploring the exhibition site, especially while interacting with the system. This way, the system is aware of the exhibits already visited by each user. In addition, users are also asked to state personal interests prior to using the AgentSalon system and they are also allowed to rate each exhibit they visit during their tour. While exploring the exhibition site, each user is accompanied by a virtual personal agent , represented visually on the mobile device as a simple, static comic figure. The technology behind these mobile virtual characters will be explained in detail in Section 3.6.2.

During the exploration phase, users of the system are free to visit a stationary information kiosk. Once in front of the information kiosk, the user's virtual personal agent is able to migrate from the user's personal mobile device to the information kiosk (see figure 3.26). The migration is done in a very simple way, by turning of the characters visual representation on the mobile device and by showing the same visual character representation on the information kiosk. The information kiosk allows for direct user interaction with the system by integrating a touch-screen device. In this way, a single user can request additional information regarding the exhibition.

However, the main purpose of the information kiosk is to foster direct communication between several users of the system. The system allows up to five users to benefit of a single stationary device. Each of these users' virtual personal agent will transit to the information kiosk, as described above. Since each user chooses a different visual representation for his/her virtual personal agent, these characters may all appear on a single device and still allow users to easily distinguish the different characters. Once a virtual personal agent appears on the information kiosk, the AgentSalon system detects common as well as differing parts in the different users interests and visiting

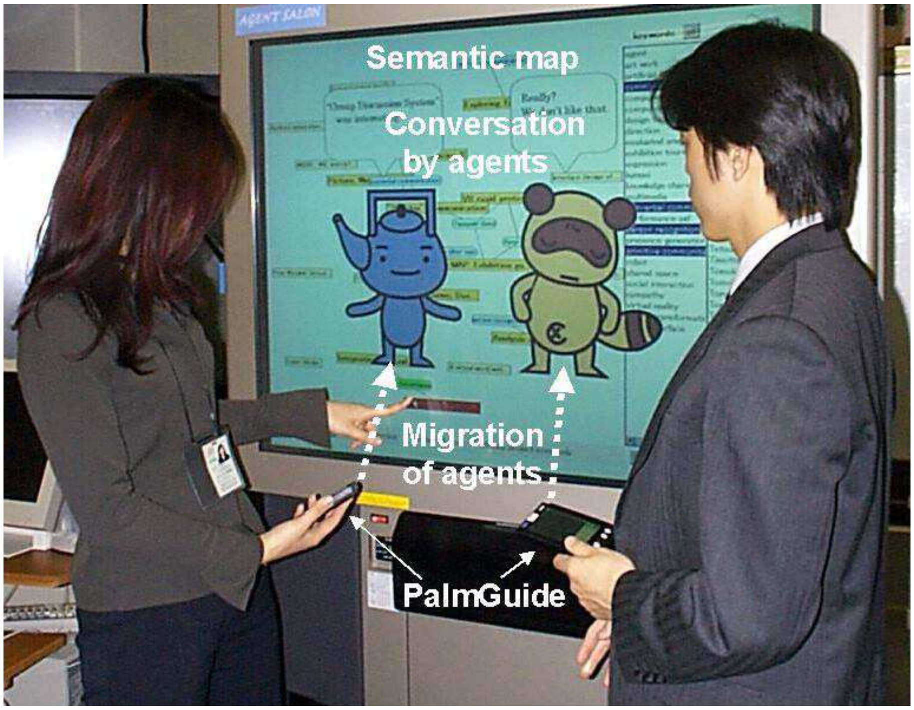


Figure 3.26: AgentSalon: Multi-user and multi-device interaction (source: Sumi & Mase, 2004)

records. Based on a calculated overlap between user interests and experiences, the agents plan and begin inter-agent conversations observed by the users. The dynamic script-generation for interesting inter-agent communications is based on a knowledge-based system. Using strategic rules, the system combines reusable templates of conversations, resulting in a conversation script for several virtual personal agents.

The basic idea behind these inter-agent communications is to start a communication between their users (as illustrated in figure 3.26). While the personal agents talk about the interests and experiences of their users, hopefully users will be encouraged to start a conversation among themselves about the suggested topics.

3.5.8 Summary

The projects reviewed in this section give an overview on the many different ways in which combinations of devices may be used in order to support new functionalities. By combining devices from different classes (i.e. mobile and stationary devices), it is possible to overcome the shortcomings of each device class. While mobile devices offer the benefit of being available at any time and any place, the processing capabilities as well as the presentation quality cannot match the one of larger, stationary devices. While some of the discussed projects use a device combination in a parallel mode (i.e. using the devices together at the same time), others make use of them sequentially. The sequential use is of special interest in application scenarios combining both mobile and stationary components. Depending on the given situation, either one of the device classes is used.

When combining devices in parallel mode, the resulting device combination offers the potential to explore new interaction metaphors. A mobile device may, for example, be used as a remote control or it may serve as a tool to access multimedia content on a stationary device. However, most of the given examples, when using the devices in parallel mode, assign a different purpose to each of the devices. The only counter example is the CommChair project (see Section 3.5.3) in which a number of devices are used in an equitable fashion by combining the displays of the devices in order to form a display continuum. The combined devices in the CommChair project do not only offer a large display continuum, but the way in which the devices are combined also results in a new interaction metaphor. Data located on one device may be transferred to another by simply dragging a visual data representation on a source device and dropping it on a destination device.

None of the reviewed projects however explores possible solutions for combining mobile and stationary systems in a parallel mode as the one mentioned in the CommChair project. Instead of using a combination of devices of a single device class in order to form a display continuum or a complex presentation system, a combination of devices coming from different device classes could effectively combine the advantages of both device classes. For example, when presenting information on a public, stationary system, a mobile device could be used to supplement this information with private information shown on the mobile device in parallel (as discussed in Section 4.5 as well as in chapter 7).

3.6 Life-Like Characters in Mobile Applications

In this section we will review projects which realized life-like characters on mobile devices. The characters described in this section are far less complex (both with respect to their visual appearance as well as their behavior) than the ones described in Section 3.1. This is mainly due to the fact that a character engine running on a mobile device has to deal with the inherent limitations of mobile devices. These limitations (e.g. small screen size, low computing power) have an obvious impact on the visual appearance of the resulting character. While the screen size limits the size of the character (especially when the screen space is also used to show additional information), the limited computing power of mobile devices has a direct influence on the possible complexity of the visual appearance of the life-like character. However, the visual appearance is not the only aspect of a life-like character which is influenced by the fact that it is running on a low power mobile device. Communication between characters and users is also limited as compared to characters running on stationary systems. While a character on a modern stationary computer may use natural language input and output as an interaction metaphor, the quality of such services on a mobile device is far lower which results in a bad recognition quality of spoken input as well as low quality synthesized speech for the character.

One way to overcome the limitations caused by the low computing power of mobile devices is to integrate additional stationary devices in the application scenarios. The complex computational tasks, for example, speech recognition and speech synthesis may in this case be transferred to a powerful stationary system. The mobile devices in this case are solely used to gather the raw input data and to render the results generated by the stationary devices. An example of such an approach is the SmartKom-Mobile project (see Sections 3.5.6 and 3.6.1).

A second solution in order to deal with the limitations of mobile devices when realizing life-like character engines is to reduce the character's capabilities until it fits the capacity of the mobile device. Examples of such minimalist approaches are the AgentSalon and PalmGuide projects (described in Sections 3.5.7 and 3.6.2).

3.6.1 SmartKom-Mobile

SmartKom-Mobile (see Malaka, Haeussler, & Aras, 2004; Bühler, Häussler, Krüger, & Minker, 2002) is the mobile device based version of the SmartKom system described in Section 3.5.6. As described above, the SmartKom architecture consists of up to 40 components. The complexity results in a resource problem on the mobile device due to both limited computing power and memory resources. In order to solve this problem, the SmartKom-Mobile prototype relies on two high-end PCs which are connected with the mobile device via a wireless network connection. Even though this setup may be impractical for a real-life use of the system, the SmartKom project members envision a future in which the computational power and knowledge bases necessary for the system to run, will become accessible through new wireless broadband networks such as third generation cellular connections like UMTS (Universal Mobile Telecommunications System) or wireless local area networks (WLAN). Instead of simply porting the SmartKom prototype onto a mobile platform, the SmartKom-Mobile prototype realizes ubiquitous access to the SmartKom services through any single device and also through multiple devices. The prototype supports two different hardware setups: inside a car (using the display of the car navigation system, a microphone array for free speech input and the built-in car audio system) and a PDA-based version for a pedestrian navigation scenario (see Figure 3.27). Since both scenarios are navigation applications, SmartKom-Mobile integrates geographic information systems (e.g. GPS) into the SmartKom system.

The visual information shown on the mobile device is exported by one of the stationary PCs running the SmartKom system. A section of this PC's screen (this section has exactly the resolution of the mobile device) is grabbed and send via wireless LAN, using the Virtual Network Computing protocol (VNC[5]), to the mobile device, where the captured image stream is rendered. In addition, any interaction the user takes (e.g. tap on the touch-screen of the PDA) is reported back via VNC to the stationary PC, where these events are handled, as if they had occurred locally on the stationary device. The described setup allows the SmartKom-Mobile system to run exactly the same way as the stationary versions of SmartKom. In particular, this also means that the interface agent Smartakus is also available in the mobile version. However, due to the limited screen size of the mobile devices, the visual representation of Smartakus has been reduced to its head and hands. Specific character features were kept in both mobile and stationary versions of SmartKom to allow for easy recognition by the users (compare Figure 3.27 and Figure 3.25).

Both SmartKom-Mobile scenarios (car and pedestrian navigation) may be used in combination. Upon entering the car, the user places the mobile device into a cradle. The SmartKom mobile system recognizes the situation automatically and activates a larger screen in the cars front panel as well as the hands free car audio system. The Smartakus Character and the SmartKom application are automatically re-routed from the mobile device to the build-in display of the car. However, as

[5] An exemplary VNC implementation is produced by the RealVNC company and can be found at: http://www.realvnc.com/

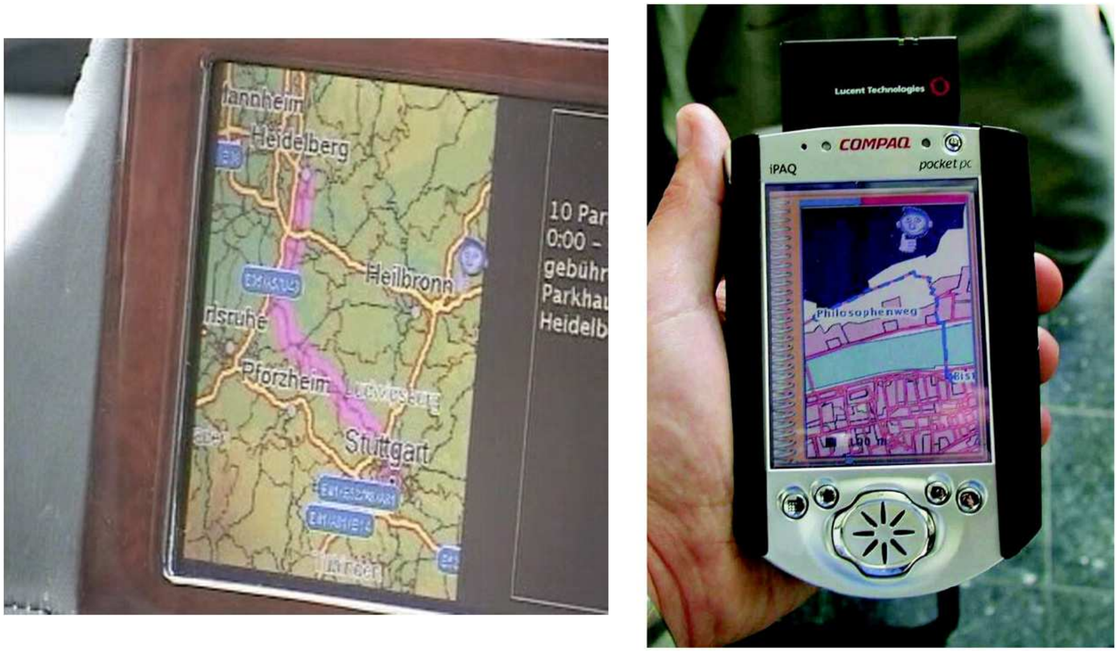


Figure 3.27: SmartKom-Mobile in a car (source: screen capture of a video clip available at Kowalski, visited 2006) and SmartKom-Mobile running on a standard PDA (source: Bühler et al., 2002)

opposed to the Migrating Characters, in neither case the Smartakus Character is executed directly on the device. Instead it is realized as a overlay stream as described above. Hence, the character transition from one device to another in SmartKom is not a character transition in the sense of the Migrating Character definition (see section 1.1).

3.6.2 AgentSalon and PalmGuide (Revisited)

While in Section 3.5.7 we gave a general overview on the PalmGuide and AgentSalon system, in this section we will highlight the life-like character technology incorporated in the system. There are two different classes of virtual personal agents which were developed within the scope of the AgentSalon and the PalmGuide projects (as described in Sumi et al., 1998; Sumi & Mase, 2001a; Mase, Sumi, & Kadobayashi, 2000).

The first class of agents is used on the stationary device, the so-called information kiosk. Since these devices are composed of standard PC components and the operating system in use is Microsoft Windows, the project developers decided to use the Microsoft Agent platform[6] for the agent implementation. Eight different agents were designed to be chosen individually by the users. Each of these agents features about 40 different kinds of actions such as pointing, greeting and moving. As mentioned above, the different agent designs are important to allow users to distinguish the different agents while they appear together on the screen of the information kiosk. While on the information kiosk, the agents perform an automatically planned communication.

[6]The Microsoft Agent platform by Microsoft Corporation may be found at http://www.microsoft.com/msagent/

The agents utterances are realized as written text in bubbles and are supported by gestures and movements performed by the agents.

The second class of agents was developed for the mobile devices used in the PalmGuide and AgentSalon scenario. Due to the limitations of these mobile devices, the agent engine described above could not be used. Apart from the limited screen size and computing power, a major limitation of the chosen mobile device (an IBM WorkPad, identical in construction with a Palm III PDA) was the operating system PalmOS. Quite obviously, this setup does not allow for the use of the Microsoft Agent platform. Instead, the virtual personal agents were implemented as a simple graphical overlay on top of the PalmGuide application (see left hand side of figure 3.28).

Figure 3.28: Virtual character overlay and user query on a PDA in PalmGuide (source: Sumi & Mase, 2001a)

Instead of animated agents, in this case the virtual personal agents are realized by different still images showing the agent in different poses. As soon as the agent becomes visible on the screen, the underlying information of PalmGuide is no longer visible, since transparency is not available in PalmOS. In addition, the agents utterances are, as in the case of the stationary system, realized again by means of written text in bubbles. This results in even bigger screen areas obscured by the agent overlay.

The user interaction with the system mainly occurs while using the mobile device. Indirect interaction is given due to the user tracking mechanism realized with infrared beacons. The mobile device offers content related to the actual location of the user. When leaving an exhibit, the user is asked to rate the exhibit. This information is then used in two different ways. Firstly, depending on the user ratings, the system suggests further exhibits to visit. And secondly, the ratings and the list of locations visited is used as a basis for the agent dialogs on the information kiosk (see also Section 3.5.7). For example, if two users have similar ratings for specific exhibits but only partially overlapping lists of already visited exhibits, the agents will make suggestions to each other to visit those exhibits missing in the list. On the other hand, if users gave very different ratings for the same exhibit, the agents will point out this fact, hoping to encourage the users to discuss their different opinions.

3.6.3 Summary

Due to the complexity of life-like character technology, realizing such characters on mobile devices like PDAs is a challenge. The two projects discussed in this section take very different approaches. In the AgentSalon and PalmGuide project, the developers decided to solely rely on the capabilities of the PDA in use. The resulting character engine is, due to the limitations of the mobile device, very simplistic, namely a still image (bitmap) of a comic character in combination with written text in a bubble. The characters realized in this project lack the capabilities for natural interaction with users by means of spoken language communication and they also do not support gestures or facial expressions. It is questionable as whether such a character yields any benefit as compared to a standard, text-based user interface. In the AgentSalon and PalmGuide project, the main purpose of the character is obviously to work as some kind of reference point for the users when using the stationary system on which several characters may appear.

In order to realize a more complex life-like character engine on a mobile device, the SmartKom-Mobile project has taken a very different approach. Instead of developing a character engine capable of running on a stand-alone mobile device, the character engine as well as all necessary services are run on additional, stationary computers. The visual representation of the character as well as the application behind are rendered on one of the stationary devices which then exports part of its display to the display of the PDA. Hence, the resulting technology has a very low demand for computational power on the mobile device. However, exporting the display from one computer to another results in a high network load which consequently limits the number of mobile devices which may be used simultaneously within the scenario. Furthermore, wireless networks offering sufficient capacities for this type of application are still only available in few locations while those networks which are available almost everywhere (e.g. mobile phone networks) do still not offer enough bandwidth.

3.7 Synopsis

Throughout this chapter we reviewed several projects which deal with life-like character technology, localized services, mobile applications in general and/or combinations of mobile and stationary devices and we have identified and discussed the deficits that exist in each of these research areas. Each of these different research areas is influential for the development of the Migrating Character technology. In the remainder of this chapter we will discuss the importance of each of the research areas for our own project. Furthermore, we will indicate which attributes of each research area will play an important role during the development of the Migrating Characters and which particular problems of each area will be addressed by the Migrating Character technology.

Before we start our discussion, we have to state that the selection of projects and systems we reviewed in this chapter represents only a subset of a large number of related research projects.

Our goal in deciding which projects to review in this chapter was to select a representative subset of projects belonging to each of the aforementioned research areas. Therefore we focused on projects which highlight important, general aspects of each of these areas.

As we have seen in Section 3.2, mobile computing applications in general have to deal with a number of problems which arise due to the mobile nature of these applications. Firstly, the performance of the devices which come to use in mobile computing generally cannot compare with the capacities of modern, stationary devices. Mobile computing devices need to be small and lightweight, furthermore they have to run for a considerable amount of time autonomously without an external power source. Hence, mobile computing devices have to rely on low-power components whose performance is noticeably lower than the one of high end stationary devices which may consume almost arbitrary amounts of energy. The desired small size and low weight of mobile computing devices further influences the choice of components to be integrated in these systems. The result of the miniaturization of mobile computing devices is that small devices have common features of tiny displays and very limited input capabilities. In many cases, the screen size is further decreased when using a virtual keyboard on the touch screen of the mobile device as an input modality. The general limitations of mobile devices described in this paragraph already indicate that traditional human-computer-interaction techniques cannot be applied in an efficient way to mobile devices. Using the standard combination of mouse and keyboard in a mobile scenario is often impractical, sometimes even impossible (e.g. when the user is walking while interacting with the device). Usually mobile devices imitate this input device combination by using a touch screen for a mouse and a virtual keyboard (as mentioned above). However, in many situations it is very disadvantageous to have the user looking at the mobile device while interacting with it. Furthermore it may even be dangerous, for example, when interacting with a mobile device while driving a car. In such situations, a hands free interaction method which does not require a user to focus his/her visual attention on the mobile device would be a better solution.

A quite obvious solution to this problem would be to support natural spoken language for both input and output on mobile devices. Due to the limitations of mobile devices mentioned in the preceding paragraph, the realization of such services on mobile platforms is still almost impossible as long as the developers rely on a standalone service running on the low-power components of a mobile device. A possible solution is to relay complicated computational tasks from the mobile device to a stationary server. In this way, even computing intensive services could be indirectly realized on mobile platforms. Using such a service relay solution demands a constant, high bandwidth communication between the mobile device and the stationary server. This, however, is another general problem with which most mobile applications have to deal, namely the unreliability of communication channels and connections to services. Due to the nature of mobile applications, mobile devices usually have to rely on wireless technology in order to communicate with other devices and services. However, not every wireless network or connection is available at any given location. Furthermore, even if a specific area is covered by a wireless network, this does not necessarily mean that a mobile device will be able to use the network, since wireless connections are usually very sensible to environmental setups, like buildings, structures or even other wireless networks which may interfere with each other. It is hence necessary to integrate back-up mechanisms in mobile applications which ensure that the system may continue to work, even with limited functionality, in case the connection to the network and therefore to important services is lost.

As already discussed in Section 2.3, more and more mobile applications offer location sensitive services. A prerequisite for this is the availability of positional data regarding the user location as well as the location of objects and possibly also positional information about other users in the environment. Similar to the problem with unreliable network connections, mobile applications have to deal with the fact that positional data derived from any type of nowadays sensors is sometimes completely unavailable and if it is available, the precision of the positional data may vary to a large degree. Once again, a mobile service needs to address this problem by providing a mechanism to deal with imprecise or unavailable positional data. A possible solution to this problem is to integrate the users more reliable senses into the system by asking the user to help in determining the actual position (as discussed in detail in section 2.5).

The Migrating Characters described in this work try to address all of the above stated problems. The main reason for using life-like characters on mobile devices is to provide a natural interaction metaphor for mobile applications which demands as little as possible of a users attention. The Migrating Characters support natural spoken language communication as well as gestures and facial expressions in order to clarify communicative acts. However, as discussed above, realizing such a processing intensive approach on a mobile device without relying on additional servers available via network connections is impossible with today's mobile computing devices. Hence, the approach when realizing the Migrating Characters is to out-source any processing intensive service which may not be run on the mobile device to additional, stationary servers. Yet, the integration of stationary devices into the Migrating Character scenario not only arranges for relaying of tasks from the mobile device to a stationary one, but, instead, stationary devices are also integrated as presentation devices to be used in conjunction with the mobile device. In doing so, the Migrating Characters technology addresses the problem of limitations posed by the small screen size of mobile devices as well as the generally inferior multimedia capabilities of mobile devices as compared to modern, stationary computers. Furthermore, the additionally integrated stationary presentation systems (which may, for example, be a wall-mounted display or a steerable projector) may also be used to clarify references to physical objects performed by the Migrating Characters (which will be discussed in detail in section 4.2.1). Instead of simply verbally describing the object, the Migrating Character is capable of positioning itself right next to the physical object and pointing at it, much like humans do. This is realized by allowing the Migrating Characters to appear on both mobile and stationary devices and to transit from one device to another automatically. Apart from clarifying a reference to a physical object, this approach also reduces the need for very precise user positioning data. By integrating an external source of reference (the Migrating Character which locates itself right next to the relevant physical object), it is only necessary to ensure that the user follows the characters movements, but it is not necessary to know the user's exact location and orientation.

Generally speaking, the Migrating Characters are used as a coherent interaction metaphor in a scenario where an arbitrary number of different stationary- and mobile systems may be integrated. The Migrating Characters ensure a consistent behavior in any device constellation, and furthermore they are also used as a center of reference for the users in a multi-device scenario. Especially when performing presentations spanning several devices (see also chapter 7), the Migrating Characters provide an anchor point which allows the users to easily follow the presentations.

Seen from the side of the life-like character research area, we may subsume that research in this area is aiming at the realization of highly believable life-like characters. These characters are usually composed of two different parts, namely a behavior engine and a visual character representation. Both parts have a strong influence on the life-likeness of the resulting character. The most human like appearing character will never be perceived as a believable character, if the underlying behavior engine cannot support the human appearance of the character. On the other hand, a very simplistic visual character representation, lacking appropriate facial expressions and gestures, will not allow users to appreciate the intelligent behavior of the character. Many research projects are devoted to the development of very sophisticated character behavior engines (for example, the ones described by André et al., 1997 and also Marsella et al., 2004) and the realization of highly realistic visual character representations (see for example Kawamoto et al., 2004). All these projects share a common goal: to increase the believability of life-like characters. However, we believe that a life-like character is not only a composition of a character behavior engine and a visual character representation but it is also defined by its role. And it is usually the limited scope of a characters role which keeps it from being believable. These characters suffer from what we call the "shop assistant effect":

> *Imagine a customer buying a newspaper every morning at the same newspaper kiosk, interacting with the same shop assistant every day for around 60 seconds. After a year, he will have spend approximately 6 hours with that shop assistant, but it is very likely that he has never perceived the shop assistant as a personality.*

This happens due to the fact that the interaction between shop assistant and customer is limited/defined in time and scope by the function it fulfils. Most life-like characters fulfill a role similar to the one of the shop assistant. They offer access to specific information, they help while interacting with a program or a web site. As Barbara Hayes-Roth argues (Hayes-Roth, 2004), this situation might be interpreted as a good chance to feign intelligent character behavior, since characters can make a competent impression even if they possess a very limited knowledge, rationality and expertise in areas other than the one related to the role the character plays. We agree with this view; however, we suggest to let a single character play many different roles. In this way, the character will leave a profound impression of competence in general as well as in each situation (based on limited knowledge regarding the actual character role).

Table 3.1 subsumes the problems which will be addressed by the Migrating Character technology. We will continue by introducing the Migrating Character concept as well as general design guidelines for Migrating Characters in the following two chapters, which form the basis for the implemented systems which we will describe in Chapter 6 as well as for the user study described in Chapter 7.

Problems and limitations of the different, relevant research areas		Solutions provided by the Migrating Character Concept: Integration of stationary devices offering additional services	Integration of stationary presentation systems installed in the environment	Support for natural interaction by means of speech, gestures and facial expressions	Support for backup strategies with limited functionality	A single Migrating Character may be used for multiple purposes	Use of device-independent presentation software and data formats	Migrating Character are a device-independent interaction metaphor	Migrating Characters represent an anchor point in multi-device presenations	Multi-User support by using mobile and stationary devices in parallel	Migrating Characters act as a guide in both physical- and virtual worlds	Migrating Characters may dislocate themselves in physical- and virtual space	Migrating Characters communicative skills help to integrate user knowledge	Use the Migrating Character as an external source of reference
Mobile computing	Limited computing power of mobile devices	⊗												
	Small screen size of most mobile devices		⊗											
	Limited interaction facilities on mobile devices	⊗	⊗	⊗										
	Unreliable connections in mobile applications				⊗									
Characters	Limited believability of lifelike characters due to the roles they play					⊗								
	Consistent visual appearance and natural language support	⊗	⊗						⊗					
Device combinations	Demands a coherent interface for all integrated devices	⊗	⊗				⊗	⊗						
	Demands a means to guide the users attentional focus between devices						⊗	⊗	⊗					
	How to deal with multiple users using a single device?		⊗				⊗	⊗	⊗	⊗				
Location-based services	How to guide users through physical environments?			⊗							⊗			
	How to guide a users attention towards physical objects?									⊗		⊗		⊗
	How to deal with imprecise positional data?										⊗	⊗	⊗	⊗

Table 3.1: Problems addressed within the Migrating Character Concept

Conceptualization of the Migrating Character Technology

4 The Migrating Character Concept

Nowadays, life-like characters are a very common technology applied in many different application scenarios. As we have seen in the preceding chapter, both the visual appearance of these characters as well as the behavior-engines behind them are constantly being pushed to the next level. As a result, a state-of-the-art life-like character has both a very realistic visual appearance as well as a sophisticated behavior. Furthermore, a modern character will also support natural communication by means of spoken language in combination with gestures and facial expressions. Due to the high interest in this research area, even complete toolkits for developing life-like characters have been made available (e.g. Galatea, see Kawamoto et al., 2002). The developed technologies in recent life-like character projects are however not only very advanced, but from a technological point of view also quite demanding. Until now, there have been only few approaches towards realizing life-like characters in mobile scenarios, mainly because of the limitations that mobile devices pose due to their low computing power and limited interaction possibilities.

The Migrating Characters described in this work represent a new class of life-like characters being neither limited to mobile nor to stationary devices. Instead, the idea behind the Migrating Character technology is that of a beneficial combination of both classes of devices. A Migrating Character is capable of working on both classes of devices and furthermore may freely move between different devices.

In this chapter we will present the Migrating Character concept by discussing the novel possibilities when taking the step from life-like characters to Migrating Characters. In addition, we will also identify prerequisites which need to be met in order to allow the Migrating Characters to realize specific behaviors or features.

4.1 Enhancing the Flexibility of Life-Like Characters

Instead of being fixed to a single device or program or web site, life-like characters should be capable of following the users wherever they go. In this way, a character could support a user in many different situations, conveying the impression of a generally competent assistant. Furthermore, due to the many different possible situations where the character might help (i.e. a city tour guide, a museum guide, a shopping application), there is a very high potential of adequate user adaptation over a long period of time.

4.1.1 Active and Passive Character Locomotion

Character locomotion is the key element within the Migrating Character concept. It allows life-like characters not only to accompany users while exploring the physical world, but also to assist users by means of deictic gestures, highlighting both physical and virtual objects (i.e. images or symbols on a display).

We distinguish between active and passive character locomotion. In the active locomotion category we find all the methods allowing the character to dislocate itself, regardless of the user's movements, for example when the character ”jumps” from one device (see also Section 6.1 and Kruppa, Krüger, Rocchi, Stock, & Zancanaro, 2003) to another or moves along a projection surface by using a steerable projector to move the visual representation of the character (as described in detail in Section 6.2 as well as in Kruppa, Spassova, & Schmitz, 2005). Whenever the character depends on the user in order to dislocate itself, we refer to this movement as passive locomotion (for example, when the character is located on a mobile device carried around by the user, as described for example in Section 6.1). Both categories yield different advantages and implications:

- Using active locomotion, the character is capable of positioning itself freely in the environment. Active locomotion can be the result of an explicit user request or an action performed by the character itself. In the first situation, the user might have decided to direct the character to another location (e.g. from a small mobile screen, where only the character's head could be displayed to a large stationary screen to support a full embodiment of the character). Locomotion in this situation can be used to guide users through an environment. Of course it has to be ensured that the character does not loose contact to the user, which requires some kind of user tracking, either in a global or local coordinate system (e.g. distance between the user and the character).

- Using passive locomotion, the character is sure to be close to the user, however it depends on the user in order to reach a certain position.

- Depending on the chosen locomotion method, the character uses either the same (passive locomotion) or a different (active locomotion) frame of reference as the user and must change its gestures/utterances accordingly. The decision for either active or passive locomotion also depends on the actual context. A character can only migrate if appropriate screen space is available in the proximity of the user.

These observations demand different character behaviors depending on the actual/available locomotion method. We assume that the character is always driven by a certain goal. In case it is necessary to move to another location in order to fulfill a specific goal, the character could either move to the new location actively (in this case it should however ensure that the user is following) or it could try to convince the user to move to the new location and hence move the character passively. In either way, the character needs to be aware of the user's movements/actions in order to react appropriately.

4.1.2 Migration of Code

As mentioned above, a Migrating Character's movements through space may either happen indirectly, by being carried around on a mobile device or directly by moving within the projection space of a single device or by moving from one device to another. While in the first two cases, the character (i.e. the software which constitutes the character) resides on the same device, in the last case it has to be transferred from one device to another. Depending on the application scenario, there are different possible solutions to this problem, which we will discuss in the remainder of this section.

We will start with a look at an application scenario in which the Migrating Characters belong to the scenario and their use is also limited to the scenario. This may for example be a museum guide system employing special Migrating Characters developed for the particular museum (as in the case of the PEACH project, see section 6.1). If, in such a scenario, all the devices which come to use do also belong to the scenario and no additional device may be brought into it by the users, then a very simple solution, which will allow a Migrating Character to transit from one device to another, would be to install all the necessary software for the character on each device within the scenario beforehand. If a Migrating Character is to migrate from one device to another, the character (i.e. the character engine/software) would stop working on one device and start working on the other (see Figure 4.1). There would be no need to transfer any data regarding the character.

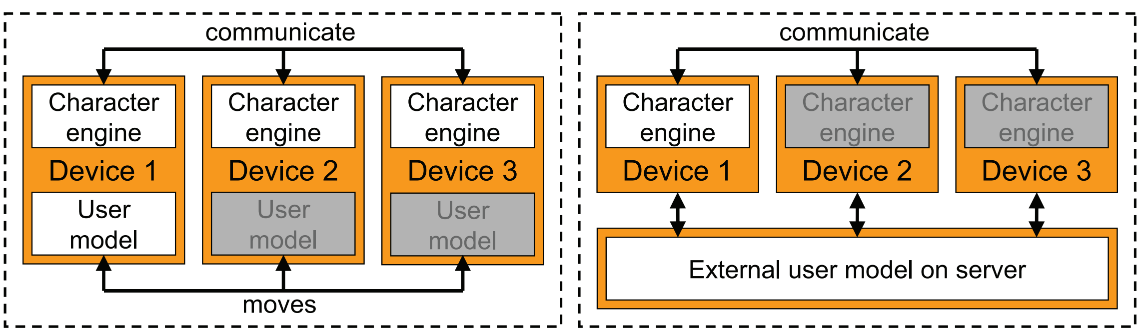


Figure 4.1: Different solutions for character engine migration between devices

However, a Migrating Character's behavior is not always completely pre-defined, instead it may adapt to the special needs of its user, based on a static user profile or by observing the user's behavior over a certain period of time. In order to allow for a consistent behavior, a Migrating

Character capable of dynamically adapting to its user's needs requires access to the user model[1] in each situation, regardless of the device the character is actually located on. Therefore, it is necessary to completely separate the user model from the character implementation itself. The user model may then either be installed and maintained at a central location (i.e. a server, see right hand side of figure 4.1) accessible by all devices included in the scenario, or it may be moved from device to device whenever necessary (as illustrated on the left hand side of figure 4.1).

There are however other scenarios, where a migration of the software constituting the Migrating Character may become necessary. This is, for example, the case when a user brings his/her own personal device, including a personal Migrating Character, into a Migrating Character installation. A Migrating Character installation may also allow a user to transfer a provided, specialized character onto his/her personal device. In either case, it will be inevitable to transfer all necessary data between the different devices, in order to realize both the visual appearance and the characters specific behavior on each device.

In Section 2.4, we have discussed the origin of life-like characters. We especially pointed out the link between the research area of agents and that of life-like characters. The problem discussed here is very similar to the one of code migration, which has been investigated intensively by a special interest group within the agent community, namely that of mobile agents. As opposed to stationary agents, the interaction radius of the mobile variants is not limited to the machine where they are started. Instead, they may migrate from one device to another, whenever necessary. This is realized by means of an architecture capable of hosting mobile agents (examples of such architectures are discussed in Baumann et al., 2002; Perdikeas et al., 1999). The agent architecture provides the environment for the agents' execution (as illustrated in Figure 4.2). Furthermore, the architecture also represents an abstraction of the underlying operating system and network.

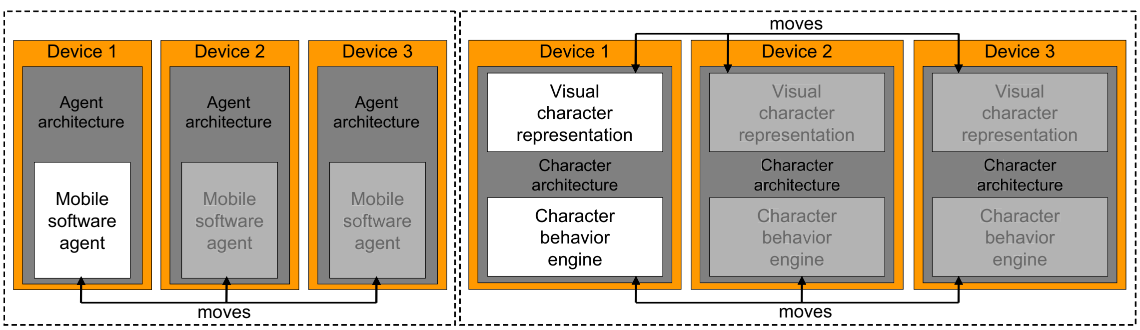


Figure 4.2: Mobile agent code migration and Migrating Character migration

There are different classes of mobile agents, going from agents which are transferred to a destination machine once and will terminate itself on that remote machine, to agents which migrate onto many different devices, taking with them both data state and execution state. Regardless of the type of agent however, the concept of migrating code is forming the basis for each of them. The mobile agent brings the code which needs to be executed on a particular machine with it, and the agent architecture on the machine allows to execute this code in a controlled way.

[1]A user model is a knowledge source in a system which contains explicit assumptions on all aspects of the user that may be relevant to the behavior of the system. These assumptions must be separable by the system from the rest of the system's knowledge (Wahlster & Kobsa, 1989)

In order to apply the concept of code migration to a Migrating Character, as opposed to an agent, it will be necessary to provide a character architecture on each device which allows to render the character's visual appearance and to apply the character's behavior engine. Hence, a clear separation of the characters visual representation and behavior engine from the surrounding application is essential. As we have already mentioned, parts of the character's behavior engine may be influenced by a user model which may come to use. The methods described above, allowing to migrate the user model from one device to another may also be applied in the same way in a scenario where code migration will be used for the character transition between devices.

4.1.3 Physical Context - Sensitivity

In mobile applications, a certain information is often only relevant at a specific physical location. This localized information, which is directly related to physical objects demands a positional awareness of the system presenting the information. Sensing the physical context is not only important when guiding a user through the physical environment, but even more when referring to a specific physical object within the environment.

The first step towards performing a reference to a physical object is sensing the object's proximity to the user. The objects could be equipped with active senders, emitting all the necessary information in a narrow range. However, it is also possible to determine the user's position and orientation, and store the information on physical objects in a database. Organizing this "world knowledge" in an hierarchical structure like an ontology will allow a system to determine relative position information (e.g. *User is in room x; User is close to object y; User can see object z*).

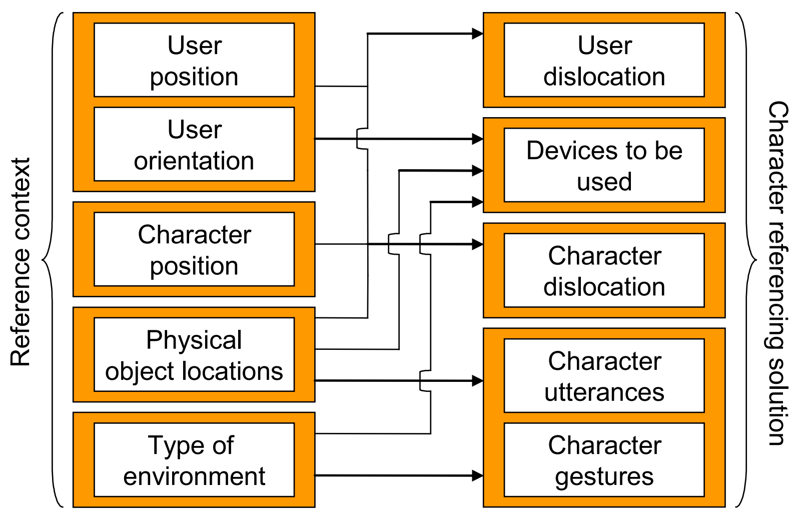


Figure 4.3: Context information influencing character reference strategy

In order to determine an optimal solution to a given referencing task within the physical world, a Migrating Character needs to be aware of the physical objects around it (i.e. the objects around the user) as well as of the physical context of the user (i.e. position and orientation). Furthermore, the character will need detailed knowledge on the object to be referenced (e.g. *What is it?; How big is* *it?; Which other objects are close to it?*). Depending on the position, size, proximity and similarity of physical objects, different strategies need to be chosen in order to disambiguate references. In Figure 4.3 we depicted the influences of the reference context (i.e the situation in which the reference occurs) on a reference plan. This graph is not necessarily complete, it rather gives a rough overview on the correlation between the reference context and the resulting behavior of the Migrating Character while performing the reference. While the user position and orientation may limit the choice of appropriate devices to be used by the character due to physical restrictions, like for example the visibility of an object, it may also indicate that a user dislocation is necessary prior to performing an object reference. References to physical objects performed by the character are based on the frame of reference of the user and the (possibly different) frame of reference of the character and the location of the physical objects. The characters gestures and spoken utterances, along with a possible character dislocation, are chosen accordingly (see section 6.3 for an exemplary realization of such an automatic mechanism to choose appropriate modalities for a reference performed by a Migrating Character).

In addition, different types of environments (i.e. a public or a private situation) may impose further restrictions on the choice of appropriate devices for a reference task. However, since users may have very different opinions on what is adequate in a specific situation, the character should also take into account the personal preferences of the user, when deciding on a specific referencing strategy.

4.1.3.1 World Knowledge

The world knowledge, necessary for the Migrating Characters to work, may be stored in many different ways. However, storing the information in such a way that the inherent structure of the physical world is preserved, may help to simplify the access of relevant information whenever necessary. A good way to preserve this structure is to represent the world knowledge in an ontology . The term ontology is borrowed from philosophy where it originally relates to the branch that studies what things exist. In the Artificial Intelligence area, the term refers to an explicit formal specification of how to represent the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them. Representing the world knowledge in an ontology allows to model the underlying concepts of the physical world objects by defining classes of objects and relationships between these classes. Once this structure is represented in the ontology, instances of the available classes of objects can be inserted into the ontology. The resulting knowledge base now allows to draw inferences over these objects. For example, depending on the definition of classes and relationships in the ontology, an instance of a chair inserted in the ontology can also be automatically identified as a piece of furniture as well as an object to sit on.

Organizing the world knowledge in an ontology will allow the Migrating Characters to draw inferences and to make useful assumptions. In this way, identifying physical objects which are located close to each other, or which are in the same room, or which fall in the same category, is very easy. An exemplary realization of a world knowledge database is described in Section 6.1.6.

4.1.3.2 Attentional Focus Model

Keeping track of the user's attentional focus is very important in mobile applications, since this knowledge may help to simplify both guiding a user and referring to physical objects. Incorporating this additional information in the planning process will allow a system to give navigational hints and physical reference clues relative to the user's momentary physical context. In this way it is possible, for example, to guide the user's attention towards a physical object by asking the user to move towards an object a few meters left of the object the user is concentrating on right now.

Unfortunately, it is not possible to measure the attentional focus. However, combining different sources of information allows to narrow down the number of objects which may be in the attentional focus of the user. The main source of information are the sensors integrated in a system. Depending on the hardware in use, it is possible to get information on both the position and orientation of the user. Depending on how the sensors are used, the orientation information may either represent the body orientation of the user, or the orientation of the user's head or simply the orientation of a device held in the user's hand. While the information on the orientation of the user's head is most valuable, it is still only a clue on the attentional focus of the user. The user may still turn his eyes towards an object which is not right in front of him. If the system is only aware of the body orientation or in the worst case of the orientation of a device held by the user, the number of possible objects which might be in the attentional focus of the user is even higher.

One way to improve the chances of successfully identifying the object which is in the attentional focus of the user is to keep an attentional focus history (see also Section 6.3). When the user is asked to focus his attention on a particular object, and the user is acting accordingly (i.e. he moves to the right spot, turns towards the right object), it is quite probable that the user is really concentrating on that particular object. Whenever the user is to move his attention towards another object, it is possible to give positional information on the target object relative to the object the user is focusing on at the moment. If the previous object is out of focus, the system can still try to determine the object with the highest probability of being in the user's attentional focus. To support this calculation, the world knowledge of the system should include information on the saliency of each object. The system may then calculate the objects which are in the estimated field of view of the user and choose the most salient object as a starting point for the referring task.

4.1.4 Personal Context - Adaptivity

In order to maximize the user's satisfaction with the Migrating Character, it is necessary to adapt to the user's specific preferences. Preferences may be related to specific devices (e.g. a user may prefer to limit the use of the PDA screen to a minimum, even though he might have to move to another room in order to do so). In addition, users should be given the opportunity to prevent the use of certain devices or device combinations in specific situations, for example a user might dislike the character to use a spatial audio system in a public situation, while the user might have a different preference in a private situation.

The Migrating Character should also be capable of determining special interests of the user by non-intrusive methods, instead of asking the user to state these interests. By analyzing the user's interactions with the system (e.g. which information does he/she request, which presentations does he/she skip or stop, how long does he/she stay in front of objects of different classes?), the special interests of the user may be estimated. However, the user should always be informed about the assumptions the character makes, based on the user's behavior. In this way, the user may understand what the character and respectively the system does and the user may also correct his/her own user model at any time.

However, since the idea of the Migrating Characters is to use them in many different situations, it is very important to put into perspective the data collected by the system prior to drawing any conclusions from the data. For example, a user visiting an art gallery may have a strong interest in the drawing techniques of the different artists while having little interest in the artworks themselves. However, when visiting a historical museum, the same user may show little interest in the technological developments in history but a strong interest in the development of art, based on the social and historical background. Hence, the data collected in a particular situation needs to analyzed in the light of the general focus of the information presented in that particular scenario.

4.2 The Migrating Character as a Center of Reference

The main purpose of the Migrating Characters is to offer a unique interaction method in complex application scenarios where several different devices and services come into use. The Migrating Character serves as a unified user interface to facilitate interaction with the system for the users. Playing the role of an interlocutor, the Migrating Character becomes the center of reference for many different mobile and multi-device applications.

4.2.1 Physical Object References

In mobile application scenarios where users are exploring physical spaces, it is of utmost importance to enable the Migrating Characters to perform unique references to physical objects.

Depending on a given situation, different reference methods may help to disambiguate physical object references. The following methods were identified to guarantee the possibility of a unique physical object reference in any given situation:

- The simplest reference method to be performed by a Migrating Character is a sole utterance with no accompanying gesture or movement (see Figure 4.4 B). However, this type of reference may only be performed, if the resulting utterance produces a unique reference.
- A more complex, but also more precise reference is achieved when the Migrating Character performs a combination of a gesture and an utterance on a mobile device. In order to refer to a physical object, this method demands a photo or abstract visual representation of the physical object, which is to be shown on the screen of the mobile device (see Figure 4.4 C).
- Another method is to let the Migrating Character disappear from the user's mobile device and reappear on a wall mounted display located close to the physical object which is to be referenced. The character may then perform a gesture towards the object and an accompanying utterance (see Figure 4.4 A).

- A fourth method is to let the Migrating Character disappear from the user's mobile device and reappear as a projection on the wall right next to the physical object which is to be referenced (in Figure 4.4 C, the character performs a reference to a painting).

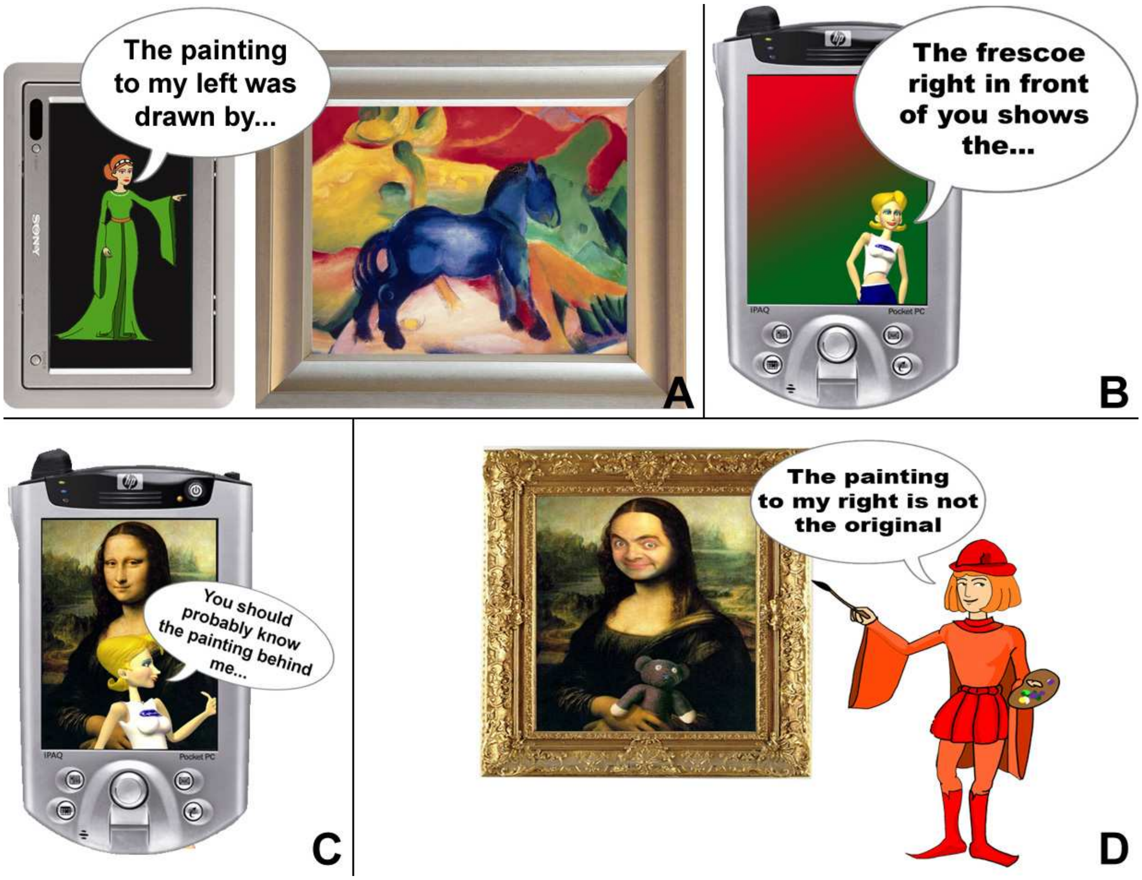


Figure 4.4: Different exemplary setups for physical object references with Migrating Characters

The Migrating Character must be able to decide automatically which referencing method to use in an arbitrary situation. This decision will be influenced by the user's physical context (i.e. the user's location and orientation), the availability of the different technical devices needed to perform each different type of reference, the personal preferences of the user and the location and class of the object to be referenced. Based on this data, the Migrating Character will identify and perform the best possible solution in order to produce a unique reference to the target object.

In Chapter 6 we will discuss implementations of each of these different referencing methods (see also Kruppa & Krüger, 2005). Furthermore, we will also present a rule-based system capable of determining an optimal referencing strategy for the Migrating Character in any given situation, by taking into account the situational context of the reference, the reference goal, as well as the user preferences.

4.2.2 User Guidance in Mobile Scenarios

Similar to stationary applications, the Migrating Characters in mobile applications may help users to navigate in the given information space. Unlike common life-like characters, the Migrating Characters are capable of giving navigational aid in both physical and virtual worlds. They may for example help users while navigating through the structured information available on a stationary screen (i.e. the characters may help to find relevant information or may give meta-information regarding information access through the system). In the very same way, the characters may offer assistance to the user when looking for a particular object or information in the physical space. Similar to the interaction on the stationary system, the character may either help the user in finding relevant information by guiding the user towards the object. Or it may help the user in understanding the underlying structure in the physical space and hence allow users to better navigate themselves.

4.3 Migrating Characters Playing Different Roles

One important goal of the Migrating Character technology is to realize life-like characters capable of assisting users in many different application scenarios, both mobile and stationary ones. However, since it is impossible to have a single character knowledge base and behavior engine which will fit in any given scenario, the character's behavior and knowledge will always be influenced by the active scenario. Due to the limitations of mobile devices, which offer very limited memory to store information and usually also only offer network access with very limited bandwidth (often there is no network available at all), the concentration of character knowledge and behavior on the actual scenario is even more important in mobile applications. On the other hand, the character should never confuse the user with a completely unpredictable behavior.

One idea to guarantee for a consistent experience with the Migrating Character is to allow the character to play different roles like an actor. Real actors may play many different roles, good actors may even let us believe in this role, but they will always put certain aspects of their personality in the role. These aspects are always noticeable, and we easily recognize the same actor (even if the actor is looking complete different) in each new role. Similar, when playing a particular role, a Migrating Character's visual appearance may slightly change (e.g. the character could wear different clothes, it could hold a typical object fitting the scenario in one hand, it could wear a suitable hat, see also figure 5.4) as well as the character's behavior, however it would still be easy to recognize the character. Since the users recognize both the character and the role the character plays, it is far more acceptable if the character's expertise is limited to the scenario as long as the character is playing the role. The idea of life-like characters playing different roles allows a Migrating Character to appear as a generally competent assistant while in each situation the knowledge and behavior of the character may be quite limited. Changing a characters role would mean to change the characters knowledge base according to the new scenario/role and to indicate this role change for the user (e.g. different clothes, assets, or the character may even mimic typical behaviors directly linked to the role it is playing). For example, a Migrating Character could play the role of a museum guide, presenting knowledge stored in the museum databases, wearing for example medieval clothes fitting the museum scenario. Upon leaving the museum and entering the museum store, the character could change its appearance, wearing more modern clothes and offering advices on the articles sold in the shop. Modeling a new role would mean to decide on the knowledge necessary to play the role and to find a mean to indicate the role the character is playing. However, even smaller behavioral changes may represent different roles. For example in the PEACH project (see section 6.1) a single character, while presenting information to the user, may either play the role of an anchorman or a presenter. While acting as an anchorman, the character announces video clips and while acting as a presenter, the character does the whole presentation on its own. In this case, the two different roles may be linked to the same knowledge database and it is not necessary to indicate when the character switches between these two roles.

Furthermore, in certain scenarios it may be interesting to use a number of different life-like characters, each playing a different role (i.e. representing a different stereotype). Since these roles (and the corresponding visual appearance and behavior) give a strong indication of what the characters area of expertise is, the users would be free to choose different experts in the same scenario and hence get many different views on the same topic (this approach has been realized in the PEACH project described in Section 6.1). We will further discuss the way in which different application scenarios may influence the design of a Migrating Character in section 5.4.

4.4 Migrating Characters Emerging From Virtual or Physical Objects

Sometimes, a Migrating Character may have a direct link to a physical object in a specific scenario. For example in a shopping scenario there may be Migrating Characters which offer information on a specific product or a range of products from the same manufacturer. Often, these products or lines of products already have some kind of character or symbol which was especially developed to be used in commercials and other advertising material (often these characters or symbols are also printed on the product itself). Since these characters and symbols are usually very well known among customers, using them in a shopping scenario is very suggesting. However, in order to clarify the link between the Migrating Character and the object (or product, in the case of a shopping scenario), the character should not appear until the user is close to the object and the way the character appears should illustrate the link between the object and the Migrating Character. A possible way to achieve this goal is to invent a method which will allow a Migrating Character to emerge from the object itself. This effect is inspired by the ancient tales of a genie in a bottle. The Migrating Character should emerge from the object like the genie emerges from its bottle (Figure 4.5 illustrates this idea). A character emerging from an object in such a way will immediately appear as a servant which is linked to the object (as is the service this character may offer).

While the solution which will allow a Migrating Character to emerge from a virtual object is very simple (i.e. it is only necessary to build an animation illustrating how the character emerges from the particular virtual object), letting a character emerge from a physical object is far more complicated. First of all, it will be necessary to use a projection either on the object itself or very close to the object. In order to do so, the position of the object must be either fixed or it must be possible to track the position of the object. Secondly, since the surfaces of most objects are not ideal projection spaces, they may only serve as a starting point for the animation illustrating the emerging of the Migrating Character from the object. During this animation, the projection should already move towards a better projection surface close to the object (like a wall or any other flat,

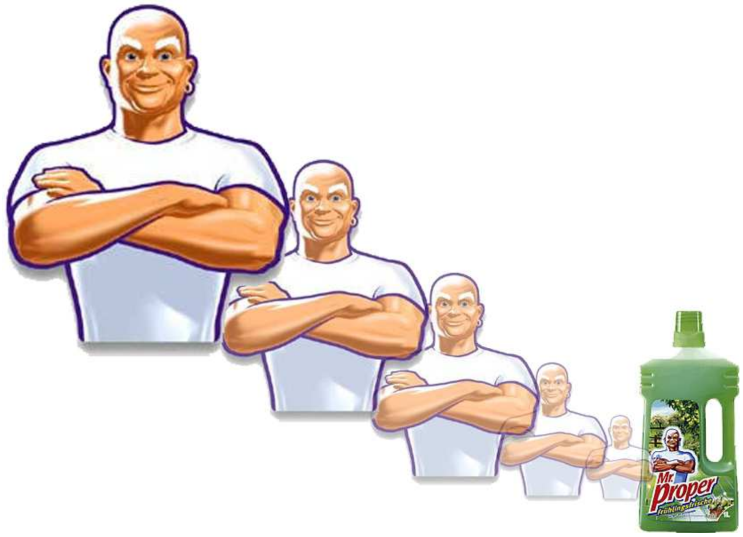


Figure 4.5: A Migrating Character emerging from a product

white surface). The benefit of a Migrating Character emerging from a virtual or physical object lies in the fact that this character as well as its purpose and knowledge is directly linked to and limited by the corresponding object is absolutely obvious and immediately understandable. An example of a Migrating Character being linked to a physical object was realized in the PEACH project and is described in Section 6.1.2.2.

4.5 Interaction With Migrating Characters in Public Scenarios

The fact that the Migrating Characters will very often be used in public scenarios demands special caution when deciding how to present information to the users and which devices to use in different situations. In the following subsections we will discuss reasons why information presentation by Migrating Characters may be sometimes problematic in public scenarios and how these problems may be solved.

4.5.1 Privacy Issues

Presenting private information in public scenarios is a commonly known problem and many different approaches have been evaluated over the last years (exemplary approaches using specialized hardware are discussed by Yerazunis & Carbone, 2001 and also by Shoemaker & Inkpen, 2001). While it is desirable to present information in the most comfortable way (e.g. on a large screen, on a high quality video and audio system), it is also necessary to keep private information out of the range of others. In a system which combines personal, mobile devices with public, stationary systems, the personal device may always be used as some kind of backup mechanism (as indicated in Figure 4.6). When no other presentation device or modality is available for presenting private information, the personal device may always be used to deliver the information, even though possibly in a lower quality.

Figure 4.6: Using the mobile device as a backup mechanism to ensure privacy

In order to determine whether a shared public information system may be used to present a specific piece of information to a particular user, first of all, the information needs to be classified as either private or public. The decision whether a certain class of information is to be considered private or public should be left to the users. Once these categories are known, new information could be automatically allocated. If a specific piece of information was classified as private information, it will be necessary to determine which public devices may be used to present the given information in the best possible way. In addition, knowing whether other users are standing within the viewing and hearing range of the public device is essential. Tracking all users in a system would be one possibility to determine whether a public device is safe to be used for private information display. However, users should also always be asked, whether they prefer to use the public or the private display for the specific information.

4.5.2 Social Aspects

In addition to the privacy issues discussed above, social aspects play a very important role when interacting with computer-based systems in public scenarios. Different users may have very different preferences regarding the interaction modalities to be used in different situations. This is comparable with the use of mobile phones. Some people don't mind at all, when talking with a loud voice on their mobile phone in the middle of a big crowd. Others would never dare to use their mobile phone in such a situation and feel absolutely ashamed when their phone rings in a public situation.

Migrating Characters allow to use many different modalities for human computer interaction. Since these characters imitate human beings, the tendency is to use natural voice input and output. In a private scenario, this may be adequate. However, many users may feel uncomfortable in case they should be forced to talk to a life-like character in a public scenario. The situation may become even worse in crowded places with a lot of background noise, where users would have to speak very loud and to articulate clearly. The same holds for the utterances the character produces. Some users may be fascinated by the technology, regardless of the other people around them, others may feel that they might disturb somebody when the character starts talking to them.

There are many ways to change the situation. For example, using a headset would completely eliminate the problem with the characters utterances possibly disturbing others. This headset could be connected to a private device and would hence not limit the mobility of the user. The microphone on the headset could also improve the speech recognition quality while also allowing users to speak in a lower voice.

However, this is still not a solution for everybody. Many people will simply not like to talk to a machine in a public place at all. So it should also be able to eliminate this modality completely. This goal may be achieved by exchanging the speech output of the character with written text and using traditional input methods like a mouse or a touch screen and a keyboard. In addition, the private device of a user could be used as a remote control for a public device. This would allow the user to benefit of the possibly larger screen and better audio system of the public device, while it would also allow the user to interact with the public device untroubled.

4.6 Synopsis

The main goal of this chapter was to introduce the conception of the Migrating Character technology. As we have discussed, the Migrating Character concept does not only promise to bring new possibilities for life-like characters, but it also means that these characters, by entering a completely new application realm, will face problems and limitations which were unknown to previous life-like character projects. As the most important underlying idea of the Migrating Characters, we identified the characters ability to relocate themselves in both physical and virtual worlds autonomously. In this context, we introduced two different types of character locomotion, namely passive and active character locomotion. While passive locomotion refers to situations in which the characters movements are based on the user's locomotion, active locomotion means that the character moves from one place to another without the need of a user locomotion or interaction. It is important to realize both types of locomotion when implementing Migrating Characters which should be both capable of following a user in physical environments but also to guide a user through the physical world. We continued by identifying important information, like the physical context and world knowledge in general as well as a model of the user's attentional focus, which becomes necessary when a life-like character is amplified with the ability of locomotion in physical environments. We especially stressed the importance of this contextual information in situations in which a Migrating Character performs a reference to a physical object.

Equally important in mobile applications however are the user's personal preferences. When trying to determine the best solution in order to guide a user to a specific location or to refer to a physical object, a Migrating Character should also take into account the user's preferences. These preferences may be both related to devices only or they may also restrict the use of a particular technology at a certain location or in a specific social setting.

We continued by discussing the way in which the Migrating Character may be used as a center of reference in mobile and multi-device applications. Since the idea of the Migrating Characters is not only to allow them to move between stationary and mobile devices, but also to allow for a high degree of flexibility of each character by using them in many different scenarios, we discussed the idea of letting a single character play many different roles, each fitting a particular scenario. Much like an actor, the Migrating Character would change its appearance as well as its behavior slightly to adapt to a new scenario, however it is always necessary to ensure that the character is recognized as always being the same.

Sometimes however, a Migrating Character may be developed for a special purpose, being linked to a specific physical object. These characters, as opposed to the Migrating Characters which are constantly guiding a user through the environment, should appear only when a user steps close to the object they are referring to. In order to further illustrate the link between this special character and the physical object, the character could emerge from the physical object itself.

We concluded the chapter by discussing implications for the Migrating Character technology due to the fact that these characters will be used in private but equally often in public settings. In this context, we discussed both the importance of privacy issues as well as the need to take into account social aspects when planning the behavior of a Migrating Character while being in a public setting. A particular class of problems we did not discuss in this chapter are the implications for the actual character design of the Migrating Characters. Since these characters will be used on many different devices and platforms, it is both necessary to ensure that the underlying technology is device-independent and that the character layout is suitable for both mobile and stationary devices. In this context, it may even become necessary to allow for the migration of program code from one device to another. The migrating code would allow a Migrating Character to appear on a completely new device in the scenario without installing any data regarding the character. Instead, all necessary data for the character would be transferred from the device the character is located on at the moment to the target device. Such a mechanism would demand a basic system to be installed on the new target device, but the use of different characters on the device would be completely flexible. This is especially important in scenarios, where users would bring their own Migrating Characters into a Migrating Character installation.

Each Migrating Character, as long as its goal is to be used in a flexible way on both mobile and stationary devices, must comply with certain design criteria. This means not only to ensure that the complexity of the character allows to use it also on a low power mobile device, but it is equally important to find a layout which uses large screens of stationary devices as well as small screens of mobile devices in an optimal way. In the following chapter we will discuss possible solutions to meet these requirements. We will also further investigate how to ensure coherency of the character's appearance in mixed device scenarios.

5 Design Principles for Migrating Characters

In this chapter we will discuss the several design criteria for developing Migrating Characters. These criteria are influenced by the devices used as well as the application scenarios themselves. While the inherent limitations of each device used to render the Migrating Characters may limit the complexity of the character's visual appearance, it may also influence the decision, whether a full body visual representation of the Migrating Character should be chosen or maybe just the head or the upper body of the character. Also, Migrating Characters should indicate, by means of subtle changes in their visual appearance, the type of application scenario they are operating in at the moment. In the following subsections we will discuss possible limitations for the visual character design for specific classes of devices, the implications for the character design in multi device applications and the way the application scenario may influence the appearance of the character. Finally, different ways of character movement in virtual worlds as well as in the real physical environment are discussed.

The characters shown in the figures of this chapter were realized using Poser 6 by Curious Labs[1] and were especially generated for this chapter to illustrate the presented concepts´(apart from the character in figure 5.3, which was developed for the German Research Center for Artificial Intelligence and later used in the PEACH project, see section 6.1).

[1]http://www.e-frontier.com/

5.1 Character Layout for Large Display Devices

Modern, stationary presentation systems usually offer a great potential for multi media presentations. These devices (which actually quite often are “off the shelf” PC components but may also be specialized products like the steerable projector unit used for the Virtual Room Inhabitant described in section 6.2) combine high resolution, large scale displays with a vast amount of computational power. Furthermore, they usually also offer high quality audio systems. From a technical point of view, these systems are capable of dealing with any type of character visualization technology (e.g. vector graphics, bitmap animations or even real time rendered 3D characters). However, one has to keep in mind that the Migrating Character will not be the only component demanding computational power on the device. The benefit of these multimedia capable devices is that they allow for high quality, complex presentations including video clips, audio clips, real time animations, interactive 3D animations etc. Hence, the character visualization must leave enough computational power for all the other multimedia components to be shown on the stationary system. A Migrating Character used on a large display devices (especially when used on a display continuum generated by a steerable projector, see section 6.2) has to deal with another problem. When moving from one location to another, the character may walk. However, in order to keep up the impression of a natural character behavior the character may only walk at a speed which fits its size. Hence, it may take a long time for the Migrating Character to reach its destination. Therefore, it should be possible to let the character be transported by another virtual object (see figure 5.5) or it should be transformed into a neutral object which may be moved in an arbitrary speed (see figures 5.6 and 6.6).

5.2 Character Layout for Small Display Devices

The most prominent class of devices which usually feature small screen displays is that of mobile devices. However, there are also stationary systems like, for example, electronic picture frames, with very limited screen space. In contrast to modern, stationary presentation systems, mobile devices like Personal Digital Assistants have to deal with a number of limitations regarding life-like character usage. The most obvious problem is the screen size of these devices, which has to be small in order to maintain the form factor of the whole system. Even though the resolution of these displays has been constantly growing, the physical size is still a limiting factor. In most scenarios, the life-like character on a mobile device will be used to present some kind of information and often this information will also need screen space in order to be presented properly. For example, when referring to a physical object, it may be necessary to display a photo or an abstract graphic representing the object on the display. Sometimes, it may even be necessary to be able to highlight a certain detail of such a graphical representation. A character can perform an according gesture, pointing into the direction of the particular detail. However, in order to do so, the character layout must allow for gestures. Hence, one goal when designing life-like character layouts for mobile devices with small screens is to minimize the screen space occupied by the character while still allowing the character to perform simple gestures. Figure 5.1 depicts possible character layouts for small screen devices.

While the layout on the left in Figure 5.1 uses a lot of screen space for the character, it also allows the character to move across the screen in a natural fashion (i.e. the character may walk,

Figure 5.1: Three different character layouts for mobile devices

since it is a full body character layout). However, since the resulting area in which information may be presented is so small, the benefit of a character capable of moving along the screen is very little. The character layout depicted on the right hand side of Figure 5.1 minimizes the space occupied by the character, by only showing the characters head and shoulders. However, this layout still allows the character to perform simple gestures by pointing in a certain direction.

A second limiting factor on most mobile devices is the computing power available, which is usually just a fraction of the computing power available on modern, stationary systems. Since in most applications, the main purpose will not be to render the character, but to present some kind of possibly localized information, the limited computing power must also be shared among different processes running on the same device. There will be at least an audio system running, in order to render the voice of the character. In addition, the system will probably show still images, maybe even animations and video clips. Furthermore, the positioning technology will also demand a certain portion of the computing power. All these factors limit the possible complexity of the character animation and the corresponding rendering process. One way to keep computational power needs low, while still allowing for a highly detailed character layout, is to opt for vector graphics when designing the character. However, vector based life-like characters always have the appearance of comic characters and do not reach the same level of realism as for example 3D rendered characters.

5.3 Consistent Character Design for Multi-Device Applications

In multi-device applications, combining several different types of devices like PDAs, notebooks and stationary systems, the design of the Migrating Character needs to ensure a high recognition factor of the character, regardless of the device used to render the character. This recognition factor is influenced by the visual appearance of the character, its behavior and its voice. All these factors need to be constant over the different devices. We will discuss possible solutions to realize a consistent character design for multi-device applications in the following subsections. However, we start the discussion by presenting methods which help to improve the recognition factor of the character by illustrating the transition of the character from one device to another in such a way that it supports users in following the character throughout the transition.

5.3.1 Visual Effects to Illustrate Character Transitions Between Devices

When a Migrating Character moves from one device to another, it is important to allow users to easily follow this movement. The first thing to keep in mind is that users may sometimes be unaware of the reasons, why the character decides to move to another device.

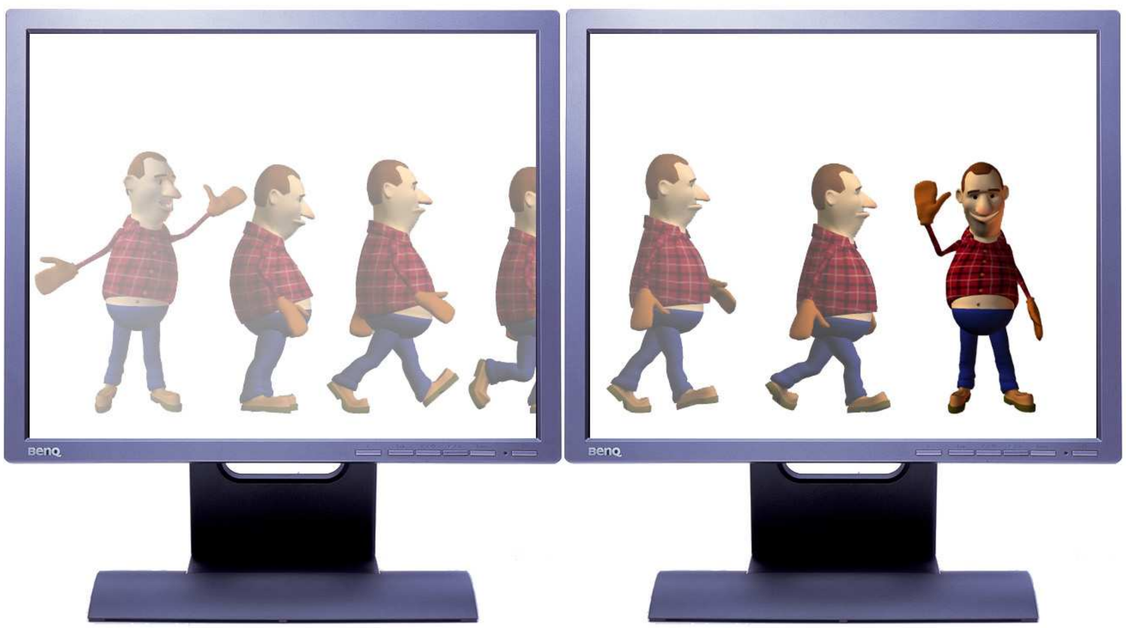


Figure 5.2: Exemplary character animation for a character transition between adjacent devices

Hence, depending on the given situation, the character should announce an upcoming device transition in order not to surprise the user and end up in a situation where the user has lost the visual contact with the character. In a situation, where the character is moving from one device to another in order to get closer to a physical object, the character could even inform the user about the fact that it will move right next to the physical object. However, in a situation where a character is constantly moving between two devices (e.g. during a presentation that spans over two different devices), the character should not announce each transition, but instead it should explain the presentation mechanism in case the user hasn't experienced such multi-device presentations before.

During the transition of the character from one device to another, accompanying visual effects might help to illustrate the transition process. Depending on the relative physical location of the two devices being used, different animations may be applied.

In the case of devices being adjacent to each other, the character animation could be a constant movement starting at the initial device and continuing on the target device. If the orientation of the two devices is horizontal, the character could simply walk out of the initial device and continue walking into the target device (as illustrated in Figure 5.2). In case the devices are oriented vertically, the character could climb from one device into the other. Due to this constant movement, it is easy to follow the character during the device transition.

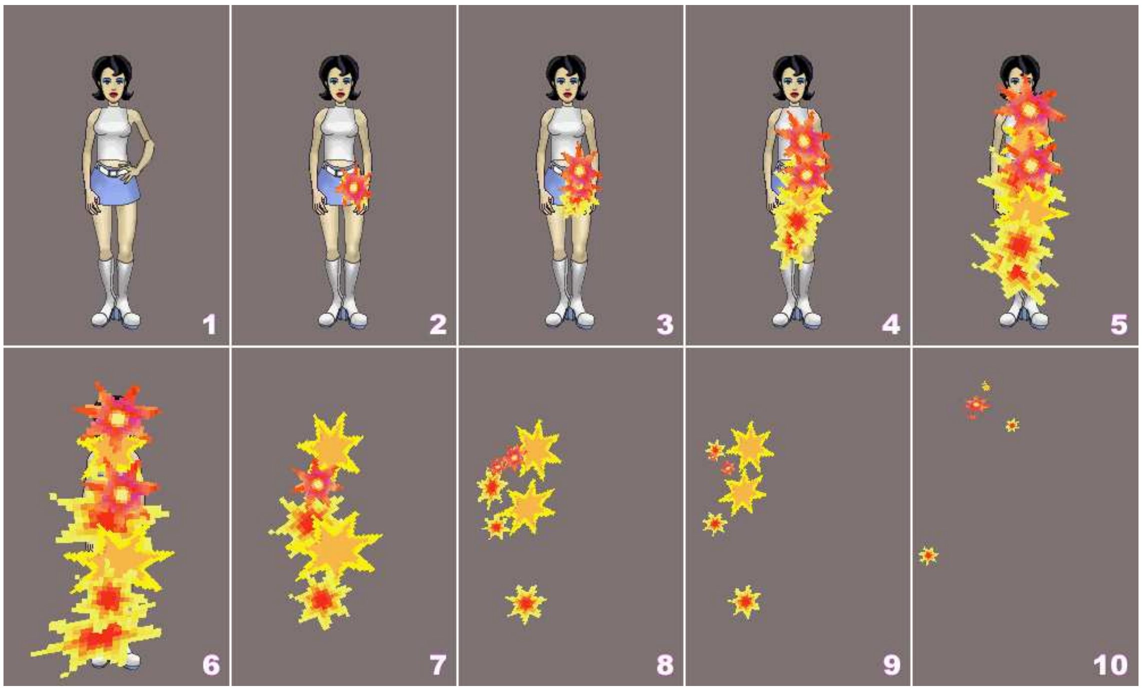


Figure 5.3: Exemplary character animation for a character transition between distant devices

If the devices are not located right next to each other, a constant character movement from one device to another is not possible. However, it is still possible to support users by illustrating the disappearance of the character on the initial device with an animation, immediately followed by an animation on the target device with an appearing character (Figure 5.3 shows an exemplary animation for a character transition between two distant devices. Played in order from frame one to ten, the character disappears from the initial device and played backwards, the character reappears on the target device).

5.3.2 Aural Cues to Support Users in Following Character Transitions

As mentioned in Section 5.3.1, a character should announce an upcoming transition to another device in certain situations. This announcement will focus the attention of the user on the character prior to starting the actual character transition and it will also help the user to follow the transition, since the character may inform the user beforehand, what the purpose of the transition is and where it will end. Furthermore, the transition itself should be accompanied by an aural signal. This signal should supplement the visual effects during the character transition. In the case of a character disappearing on one device, and reappearing on another device, the accompanying sound should also start at the initial device and move towards the target device. Depending on the device setup, this accompanying sound could make a constant movement from one device to another, in case the character is performing a constant movement (as described in Section 5.3.1) or a first sound could illustrate the disappearance of the character on the initial device, while a second sound on the target device could played while the character reappears. Either method will allow users to easily follow the character transition between devices, even if they lose sight of the visual representation of the character during the transition process.

5.3.3 Consistent Character Voices on Different Devices

A constant visual character design is very important in order to allow users to easily recognize the same Migrating Character on different devices with different visual character layouts. However, the visual appearance of a Migrating Character is not the only factor influencing uniqueness of the character in multi-device applications. A consistent character voice on different devices is also very important. If a Migrating Character would speak with different voices on different devices, the users would probably either believe that these are different characters or that the choice of different character voices on different devices would have some kind of intention behind it. To ensure that a single Migrating Character is perceived as the same character on any type of device and in any kind of scenario, we need to ensure that the character voice and quality is consistent throughout the whole application spanning different devices. One way of doing so, is to make use of pre-recorded speech, stored in an audio format which allows to be played back on all devices in a scenario in the same quality. However, pre-recorded speech would strongly limit the flexibility of a Migrating Character. If all mobile devices in a scenario are capable of using the same, high quality speech synthesis as the stationary devices, using the same speech synthesizer on each device would be sufficient. Unfortunately, the available computing power on most of today's mobile devices will limit the quality of the speech synthesizers available for these platforms. One solution could be to use a speech synthesis of lower quality which also runs on the slower mobile devices and thereby guarantee the consistency of the character voice. To suit the low power CPUs of today's mobile devices, formant-based speech synthesis is the best method since it demands low computational power and memory. Formant synthesis does not use human speech samples at runtime, but instead is based on acoustic model. A waveform of artificial speech is generated by varying parameters such as noise levels, voicing and fundamental frequency over time. The resulting speech synthesis is usually quite understandable but is also far from being perceived as natural. Using such an unnatural voice for a Migrating Character would result in a lowered overall believability of the character. The most common methods for achieving high quality, naturally sounding speech synthesis are based on the concatenation of segments of recorded speech[2]. The real-time selection and concatenation of these speech snippets, however, is very costly in terms of CPU usage and demand for memory.

[2] An overview on common speech synthesis methods can be found in (Cole et al., 1997) and also at: http://en.wikipedia.org/wiki/Speech_synthesis

A possible solution for producing a constant character voice on a very high quality level for both mobile and stationary devices is to use a client server setup for the speech synthesis. Instead of running the speech synthesis on each device, each client could send a request to a server which than would run a speech synthesis and store the result in a format playable on all devices in the scenario. The generated audio files could then be forwarded to the corresponding clients over a network. The resulting utterances of the Migrating Character would hence be of a very high quality regardless of the device the character is used on.

5.4 Character Design Influenced by Application Scenarios

While the character's visual appearance is strongly influenced by the chosen mobile and stationary devices and their capabilities (as discussed in Sections 5.1 and 5.2), it should also reflect the application scenario in which the character is coming to use. For example, in an application especially designed for young children, a rather simple comic character may be most appropriate, while a more serious character may be used for adults. The way the characters visually appear, strongly influences the expectations of the users. The comic character designed for children is probably not expected to show the same amount of expertise as a more serious character designed for adults.

Characters which are especially designed to fit in a certain scenario are expected to possess a fair amount of knowledge within the scope of the application scenario. However, due to their visual appearance fitting into the application scenario, it is also clear to the users that the knowledge of the characters is limited to the application scenario. If a single Migrating Character should be used in

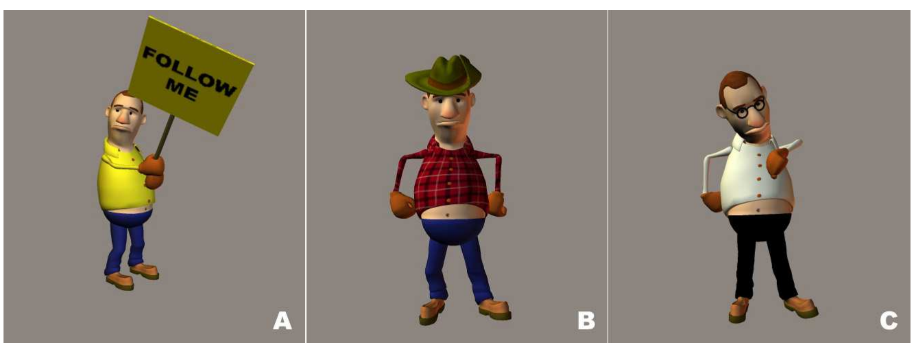


Figure 5.4: A virtual character playing different roles indicated by different artifacts and clothes

different application scenarios, it is important to convey the same sense of knowledge boundaries of the character to the users in different application scenarios. By changing the visual appearance of the Migrating Character, the link between the character and the scenario may be symbolized. However, as important as it is to allow users to easily recognize a character appearing in a different layout on different devices (as discussed in Section 5.3), it is equally important to support users in recognizing the same Migrating Character in different application scenarios. Hence, the changes in the Migrating Character's appearance have to be very subtle. Instead of changing the whole character appearance, it may often be sufficient to change a certain attribute (e.g. changing the clothes of the character, giving the character a hut to wear or putting a tool or device, fitting the application scenario, in the hands of the virtual character). Figure 5.4 shows a single character playing different roles in different scenarios. While the main characteristics of the character are constant throughout the different scenarios, different clothes and additional artifacts are used to indicate the kind of role the character is currently playing.

5.5 Character Movement

A major benefit of a Migrating Character is its flexibility regarding its physical location. Unlike traditional life-like characters, the same Migrating Character may appear at many different locations and on many different devices. However, in order to keep the believability of the Migrating Characters to a maximum, it will be necessary to allow the character to move in a natural way between different locations and devices. The transition between devices (as described in Section 5.3) is one way of dislocating a Migrating Character. But it will also be necessary to allow the Migrating Characters to move to different positions on the same device. The way this movement is realized is very dependant on the particular device the movement is performed on. As discussed in Section 5.2, a character movement on a mobile device will hardly ever be necessary, since the displays of these devices are usually so small that the characters position on the device is irrelevant.

Figure 5.5: A virtual character moving by means of a virtual transport vehicle

The situation changes, as soon as the device being used has a larger display. On a bigger display it may be necessary to move the character from one position to another in order to disambiguate a reference to an object shown on the screen. For example, when there are several similar symbols shown on the screen, referring to a particular one by using spoken language alone may be ambiguous. However, moving the character close to the object and letting the character point towards the object will disambiguate the reference. It would also be possible to highlight such a virtual object on a single screen by underlining the object or putting a virtual spotlight on it. Using the Migrating Character instead will help to provide a consistent method for object references towards both virtual and physical objects.

In order to give the character movement a natural appearance, simply moving the character (i.e. letting it float across the display) will not be a good solution. Letting the character walk instead will be the most natural solution for a character movement. However, letting a virtual character walk will not only look natural but it will also take some time for the character to reach its destination. This may be no problem while dislocating the character on a larger display. But as soon as the distance which the character will have to cover becomes too large, different solutions for moving the character from one position to another need to be found. This is especially true, when the character is not shown on a display but is instead projected on a wall with a steerable projector unit. In such a setup, the distances to be covered by the Migrating Character may become very long.

One solution is to use another virtual object to move the character. The character could for example enter a car or a plane or any other virtual representation of real objects which are commonly known to move fast. Once the character has entered the transport device, the character movement towards its destination may be realized very quickly. At its destination, the Migrating Character would leave the transport device and continue its presentation at the new location. An exemplary animation of a Migrating Character using a virtual transport vehicle to move from one location to another is depicted in Figure 5.5. Another solution to allow for a faster character movement is

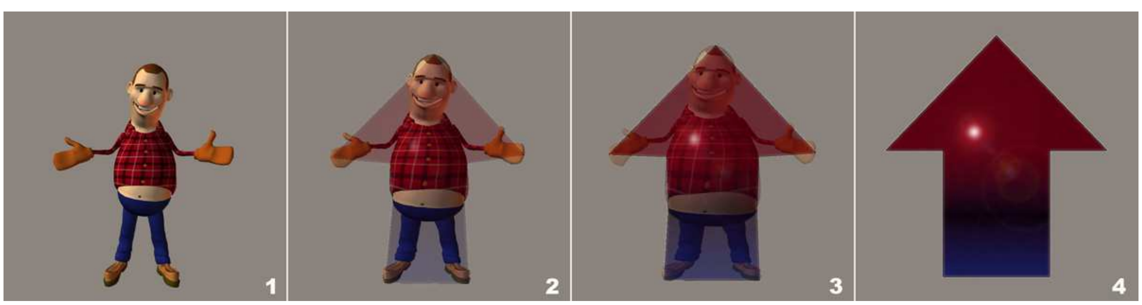


Figure 5.6: A virtual character being transformed into an arrow before moving

to transform the character into another object which may then be moved faster (see for example Figure 5.6). This may be a very abstract object, like a circle or an arrow which may then be moved without using any type of animation (i.e. the objects may float, since there is no natural movement for circles or arrows which could be expected by the users) in an arbitrary speed towards the target destination of the character. Once the target position is reached, the object could be transformed back into the Migrating Character again.

5.6 Synopsis

The main purpose of the Migrating Characters is to serve as a unique interaction metaphor in mixed, multi-device scenarios. In this chapter we have discussed issues related to the design of the Migrating Characters when taking into account the aforementioned purpose of this technology. In the first part of the chapter we discussed different character layouts to be used on mobile devices with small screens. When choosing a character layout for such a device, there is always a tradeoff between the usability of the character and the screen space it occupies. While a full body visualization of the character will allow for any type of movement and gesture, a reduced layout, for example only showing the head, shoulders and arms will reduce the flexibility of the character. On the other hand, it will also decrease the amount of screen space which needs to be allocated for the character.

Regardless of the chosen layout for the mobile device, it is of utmost importance to ensure that the character's appearance, its voice and also its behavior remain consistent when transiting to another device. This is especially true when the character is used on a completely different type of device, like for example a wall mounted, large LCD screen. Therefore, it is necessary to choose visual character representations for different types of devices, according to the capabilities and limitations of the devices. The chosen solution should always ensure that a user may easily recognize the same character in its different visual realizations. One way to support the recognition of a character when it moves from one device to another is to illustrate this transition by an animation (maybe also in combination with a sound) starting on the initial device and continuing on the target device. We have discussed several different methods to illustrate a character transition from one device to another and we identified different situations which suggest the use of either one of these methods.

As already mentioned in the preceding chapter, a single Migrating Character should be designed in such a way that it may be used in a number of different scenarios. In this chapter we discussed the use of different clothes as well as the use of additional artifacts to indicate the role the character is momentarily playing. By applying these subtle changes in the characters appearance, on the one hand a user may recognize the specific purpose of the character at the moment while realizing at the same time that this is still the same character he/she knows from different scenarios.

Finally we discussed different ways to realize a character's movement towards a specific location. These locations may either be a different display device, or a physical location or simply a different location on the same device the character is actually located at. Depending on the characters target location, it may be either necessary to let the character disappear from one device and reappear on another or the character may perform a constant movement towards its destination (in case the character stays on the same device or the target device is adjacent to the device the character is departing from). For both cases we discussed different methods. Especially in the latter case, the distance between the characters current position and its target location should influence the way in which the character moves. If the character has to cover a long distance, letting it walk to its destination might take too long and hence a different mean has to be found for moving the character to its destination. We have discussed two solutions for this problem, namely transforming the character into another object or using a virtual transport vehicle to take the character to its destination.

The described concepts and the defined principles stated in this chapter as well as in the preceding chapter form the basis for the development and realization of the Migrating Character technology in a number of different projects and applications scenarios. In the following chapter we will summarize the different Migrating Character implementations realized within the scope of this research work.

Realization of the Migrating Character Concept

6 Application Examples

As we have already discussed in Chapters 4 and 5, the Migrating Character concept was not developed for one single life-like character related project, instead it represents a general purpose concept for the use of life-like characters in mobile applications and instrumented environments. To illustrate the flexibility of the Migrating Character concept, we developed three different prototypes, each focusing on different aspects of the concept. Even though these projects do not completely cover the whole functionality indicated in the Migrating Character concept, they give a good overview on the applicability and feasibility of the concept and the technology in general.

We will start by reviewing the PEACH project in which a museum guide, based on a combination of mobile devices and stationary information systems, was realized. The main purpose of this project was to evaluate possible ways of realizing a complex life-like character on a mobile device platform. Furthermore, to overcome certain limitations of these devices (as discussed in Section 2.3), stationary information systems were integrated into the scenario. The developed character technology realizes a Migrating Character capable of presenting information in a natural way on both mobile and stationary devices. Four different characters fitting two different scenarios were developed. Each of these characters acts as a personal guide to a museum visitor and is capable of taking a user on a tour through the physical space of the museum. Furthermore, the characters also act as an anchor point for the user's attentional focus when using different presentation devices.

In a second project, the Virtual Room Inhabitant, a Migrating Character was realized which is capable of assisting a user while interacting with different devices and services within an instrumented environment. The special focus of this project was on the realization of a Migrating Character capable of moving freely along arbitrary surfaces within an instrumented room. In order to realize this technology, several devices had to be combined, most importantly a steerable projector mounted on the ceiling and a spatial audio system.

Finally, the RefFind project implements a rule-based system capable of determining optimal solutions for the task of performing references to both virtual- and physical objects within instrumented environments. It uses the technology developed in the two aforementioned projects, in order to realize a Migrating Character based referencing system to perform unique references in arbitrary situations.

6.1 PEACH - A Museum Guide Combining Personal Digital Assistants and Public Information Systems

PEACH (Personalized Experiences with Active Cultural Heritage) is a large distributed research project lead by the Center for Scientific and Technical Research at the Trentino Cultural Institute (ITC-irst). Its project partners stem from cultural heritage institutions, industry as well as from academia. A major scientific goal within the PEACH project is to design and evaluate advanced technologies that can enhance cultural heritage appreciation by creating an interactive and personalized guide. The aim is that of developing and using innovative technology to provide an educational and entertaining experience fitting the individual backgrounds, needs and interests of each user.

The German Research Center for Artificial Intelligence (DFKI), as one of the main scientific partners within the PEACH project, focused on the development of Migrating Characters to be used as virtual tour guides on mobile and stationary devices (see also Krüger & Kruppa, 2006).

Since exhibits and related information in informal learning environments are usually spread throughout the physical space, it is of utmost importance to support localized information distribution. By linking the electronically available information with a physical location, the link between physical objects and corresponding information becomes more evident. A mobile device, like a PDA, offers a high potential to present localized information in mobile scenarios in an appropriate way.

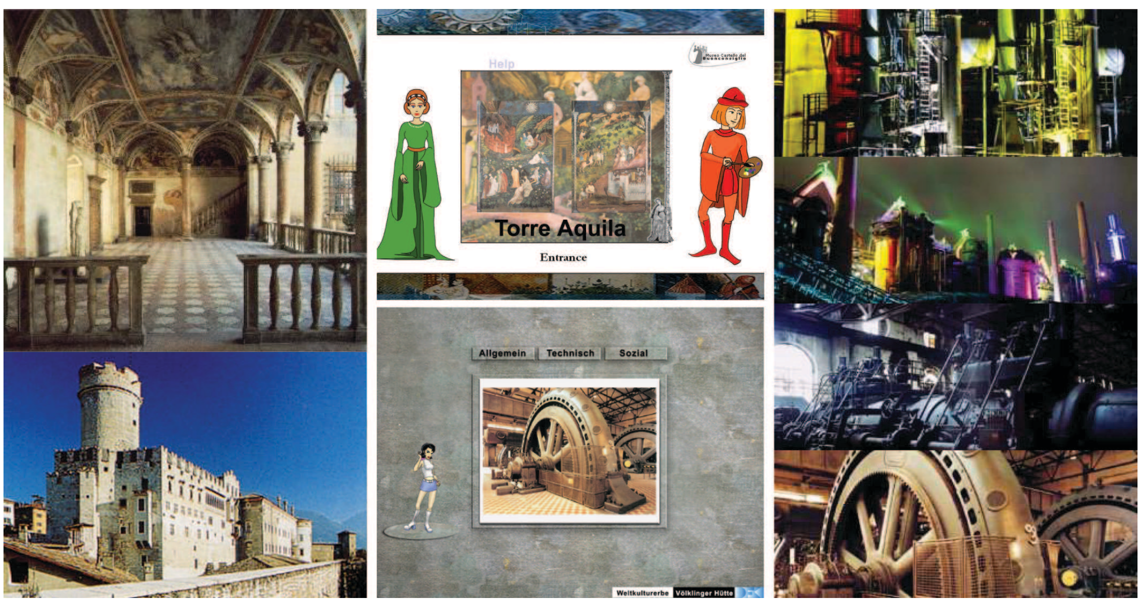


Figure 6.1: Cooperating museums and corresponding application and character layouts

However, due to the inherent limitations of such mobile devices (e.g. small display size, limited computing power), some information is preferably displayed on a large, even though stationary, information screen. In combining PDAs and large stationary information systems, the resulting information system may offer localized information as well as high quality multimedia presentations. In such a scenario, it is also very important to introduce a constant interaction metaphor, which will allow users to have a coherent experience with the system. The Migrating Characters used in the PEACH prototype are playing the role of virtual tour guides, accompanying visitors through a museum and migrating between stationary and mobile devices.

The museum guide has been adapted for two different museums, namely the Museo Castello del Buonconsiglio[1] in Trento and the Völklingen Old Ironworks[2 3]. Since these two museums have completely different settings (i.e. a medieval castle with frescoes in Trento and an industrial plant which has been recently transformed into a cultural heritage site in Völklingen), the application and character layout were modified accordingly. Figure 6.1 gives an impression of the two sites and the corresponding layouts.

The PEACH museum guide prototype combines a number of different system components. A structural overview on the setup of the prototype is depicted in 6.2. In the following sections we will explain each of the components as well as the underlying communication structure.

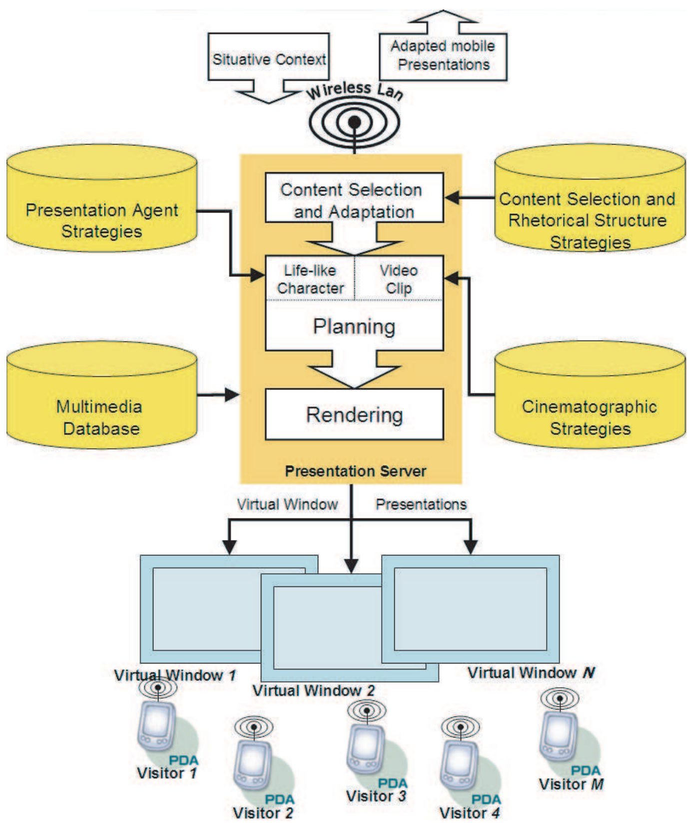


Figure 6.2: Structural overview on the PEACH museum guide prototype

[1] http://www.buonconsiglio.it/

[2] http://www.voelklinger-huette.org

[3] as part of the PIL (PEACH IsraeL) project, a third museum was recently chosen for another PEACH installation, namely the Hecht Museum (http://mushecht.haifa.ac.il/) in Haifa

6.1.1 Different Characters Representing Different Stereotypes

Four different Migrating Characters were developed for the PEACH project, two for the Völklingen Old Ironworks and two for the Museo Castello del Buonconsiglio (see figure 6.3). While the characters for the Völklingen Old Ironworks have a rather modern design, fitting the scenario of an industrial plant, the characters for the Museo Castello del Buonconsiglio have a medieval appearance fitting into the scenario of a medieval castle.

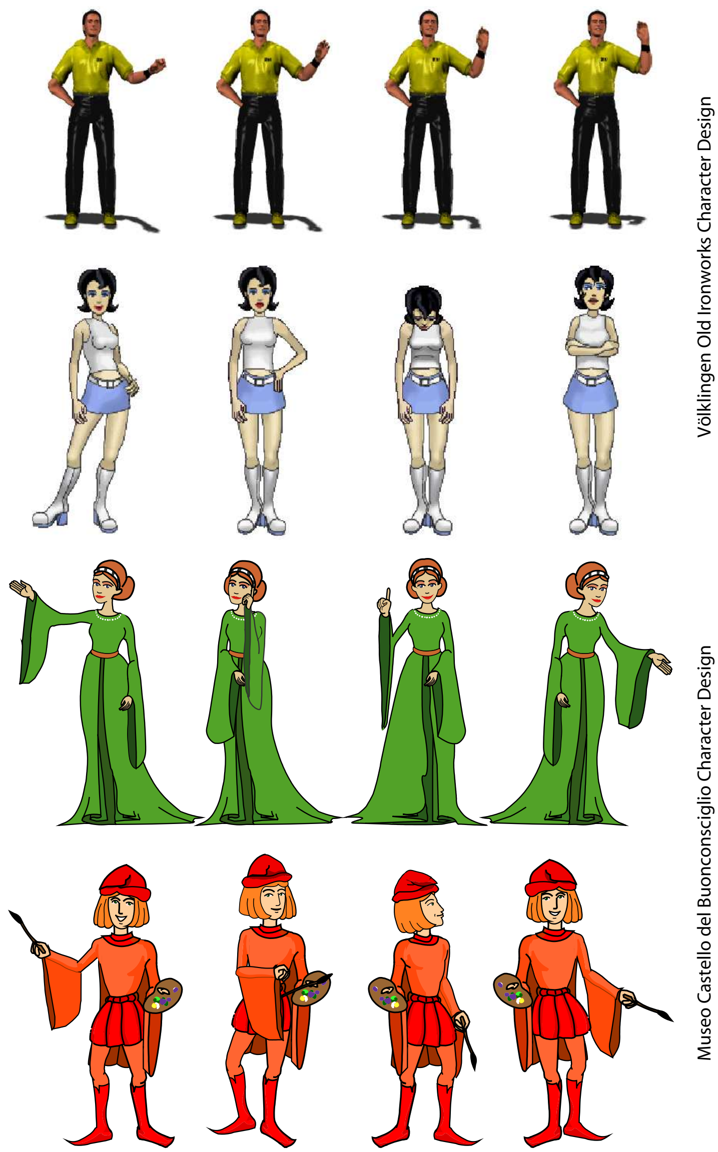


Figure 6.3: Characters designed for the PEACH museum guide

Each character represents a different view on the content of the specific museum. Users are free to choose either character at any given time. Depending on the chosen character, the focus of the information presented while exploring the museum is modified accordingly. For example, the first character in the Museo Castello del Buonconsiglio setup has the appearance of an artist, and hence the information presented by this character has a strong focus on artistic aspects of the exhibits. The design of the second character was inspired by a figure in one of the frescoes which are shown in the museum. Information presented by this second character is focused on social and historical backgrounds, since the character was designed as a contemporary witness. Both characters have easily understandable visual features indicating their role within the context of the museum. They also feature distinctive gestural behavior, which deepens the impression of the individuality of each character. Since both characters are offering the same service (i.e. guidance through the museum grounds), the interpersonal relationship is mainly influenced by this rivalry. This is reflected in the characters' behaviors by means of subtle taunts integrated in the presentation scripts.

6.1.2 Device-Independent Migrating Characters

The Migrating Characters developed within the scope of the PEACH project were designed to work on both mobile and stationary systems. Due to the different screen sizes of these devices, each character features two different layouts. While the layout for mobile devices shows only head, shoulders and arms (see Figure 6.4), the layout for stationary systems shows the whole body of the character (see Figure 6.3).

Figure 6.4: Exemplary character layout for mobile devices

Hence, the character is more flexible on the stationary device (i.e. it may move across the screen towards a specific object) while occupying less screen space on the mobile device. The characters have been realized as stand-alone movies in Macromedia Flash MX[4] and are loaded and executed during runtime in a surrounding player environment (the so-called character architecture, similar to the mobile agent architectures described in 2.4). These movies may hence be transferred from one device to another independent of the player controlling the characters.

6.1.2.1 Two Different Character Roles: Anchorman and Presenter

Following the television metaphor, two main roles have been realized within each character: the *presenter* and the *anchorman*. When playing the role of the presenter, the character introduces new media objects and uses pointing gestures. When playing the role of the anchorman, the character just introduces complex presentations without interfering with them any further. Although simpler than the presenter, the role of an anchorman can help the visitor understand many different presentations, providing a context in which they all make sense. In its role of an anchorman the character

[4] http://www.macromedia.com/software/flash/

also supports the seamless integration of the mobile devices' small screen and the large screen of the stationary system. Similar to a TV-presenter who walks around in the studio to present different content, the character is able to move between the mobile device and the stationary device (as described in Section 6.1.2.3).

6.1.2.2 Character Migration in the Physical Environment

Some characters in the PEACH museum guide have a direct link to a specific object in a particular museum. The female character designed for the Museo Castello del Buonconsiglio has a visual appearance very similar to one of the many figures in the frescoes which are shown in the museum. By giving the character this very specific appearance, it appears to belong to the fresco itself. Hence, when the character is explaining a scene depicted in one of the frescoes, it is easier to believe in the expertise of the character.

Figure 6.5: A Migrating Character emerging from a fresco

To help users to understand the link between the figure in the fresco and the Migrating Character, a simple animation was created to illustrate the character migration from the fresco to the museum guide system (as illustrated in Figure 6.5). This animation is shown during an introductory part on one of the stationary devices.

6.1.2.3 Character Migration Between Devices

In order to guide the user's attention focus while the Migrating Character is moving from one device to another, a combination of a character animation and a spatial audio clue is used. While the character disappears from the initial device, an according sound is played back on the same device. As soon as the character has completely disappeared, it starts to reappear on the target device. While reappearing, the target device plays back another sound. In this way, users are presented with both visual and aural clues which help them to direct their attentional focus towards the correct device. Figure 6.6 illustrates the character transition between different devices.

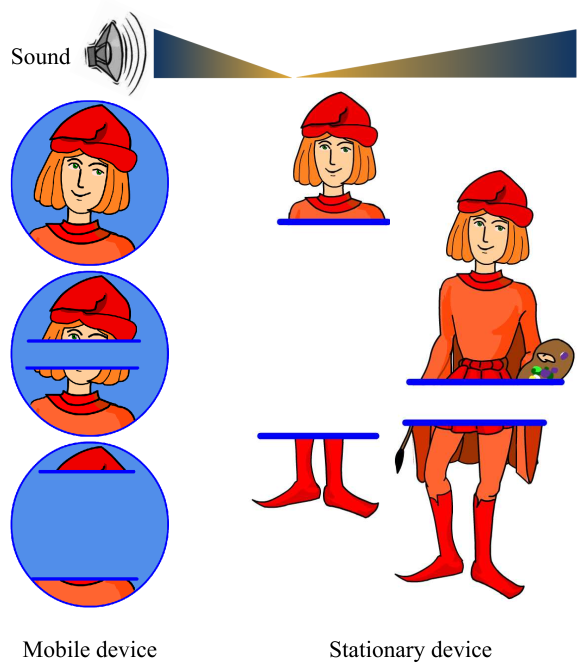


Figure 6.6: Keyframes of the transition between a mobile and a stationary device

6.1.2.4 Consistent Speech Generation for Mobile and Stationary Devices

In order to have a consistent, high quality speech synthesis available for both mobile and stationary devices, we invented a technology allowing to run the speech synthesis on a stationary server, convert the result into MP3[5] encoded audio files and share those files among all clients connected to the network. The resulting speech synthesis produces the same high quality voices on mobile and stationary devices.

We make use of AT&T's Natural Voices[6]. The quality of the generated speech of this package is very good, probably one of the best available today. To generate MP3 files from text, we use

[5] The MPEG audio layer 3 coding scheme for the compression of audio signals
[6] AT&T Natural Voices homepage: http://www.naturalvoices.att.com/

TextAloud[7] by NextUp which is offered in combination with a bundled version of AT&T's Natural Voices. TextAloud's main goal is to use speech synthesizers installed on a PC to read aloud e-mail and all other kind of text. TextAloud may also generate MP3 files instead of sending the output directly to the audio-processor. However, we use a special version, which was kindly given to us for research purposes. In this version, TextAloud monitors a certain directory and waits for text files to appear. Whenever a new text file is available, TextAloud generates a new MP3 file with same name, and deletes the text file. So, a server generates new text files in that directory and waits for the corresponding MP3 files to appear. To make sure no data is generated twice, each text file generated has a unique name. The name consists of two chars identifying the type of text and another 18 chars (being part of an MD5-Checksum[8] of the text to be synthesized), for example: S10D1Xb50cef50d41e20.txt. Before a text file is generated, the server checks, whether there exists an MP3 file with the same name as the new text file. In this way, texts already synthesized before are not synthesized again. By generating speech instead of using prerecorded speech, the system is capable of generating personalized presentations (e.g. use the name of the user, or the time of day). It is also possible to generate presentations according to the special interests of the user by combining text-parts into new text blocks. Another advantage is the easy support for different languages. Instead of translating the texts and having them spoken in a professional studio, one may simply translate them and the system does the rest. AT&T Natural Voices supports many languages (e.g. English, German, Italian, French...) and different speakers. In the demo setup at the Völklingen Old Ironworks, we have a complete support for both German and English (i.e. layout and spoken text) while in the Museo Castello del Buonconsiglio we support English and Italian.

6.1.3 Localized Information Presentation

While exploring the museum site, users are presented with localized information on their personal PDA. This information is directly related to the physical location of the user and the surrounding objects. In order to determine the user position, we use Infrared Beacons (see Figure 6.16) which are mounted on either walls or ceilings. These beacons emit a unique number and have a signal range of 8 to 20 meters and a variable transmission angle between 10 and 25 degrees. The resulting area in which the signal of an Infrared Beacon may be received is a cone.

Due to the fact that a direct line of sight between the Infrared Beacon and the receiving PDA is necessary, the system also provides information about the orientation of the user (i.e. information on the orientation of the PDA in the user's hands). To further improve the precision of the positional information, it is possible to mount several Infrared Beacons in such a way, that their emitting ranges overlap. Since the PDA can receive signals of several Infrared Beacons at the same time, we can narrow down the user's position by calculating the overlapping Infrared Beacon signal regions.

[7] The TextAloud homepage: http://www.nextup.com/

[8] The MD5 message digest algorithm, for details please see the MIT online memo number 1321, available at: http://theory.lcs.mit.edu/∼rivest/Rivest-MD5.txt

6.1.4 Stationary Information Systems

The stationary information systems within the PEACH scenario were used with the metaphor of a “window to the outside world” in mind. The stationary systems were therefore called “Virtual Windows”. While the localized information on the PDAs is directly related to a physical object, the information presented on the stationary information system offers additional information on exhibits visited while exploring the physical site of the museum. The additional material is automatically pre-selected, based on a user model which is built up, while the user explores the museum (see Section 6.1.5).

In addition, the stationary systems provide options to change the character which is accompanying the user on the mobile device. In the Museo Castello del Buonconsiglio installation, both characters present information on the stationary system (Figure 6.7 shows two screen shots of a presentation running on the stationary system).

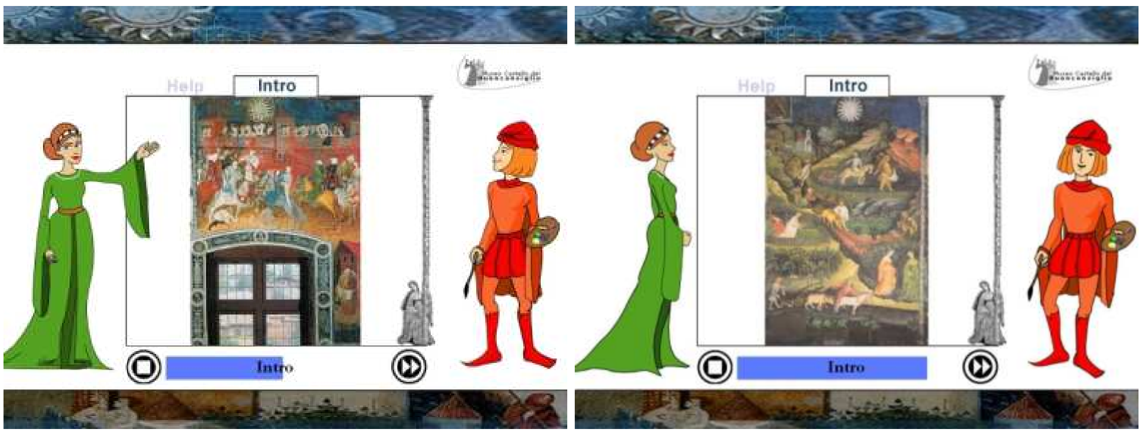


Figure 6.7: Screen shots of the stationary presentation system at Museo Castello del Buonconsiglio

However, when leaving the stationary system, the user has to decide to take either one of these characters with him on the mobile device. In case the user would like to switch to the other character later on, it is possible to choose a different character by visiting the next available stationary system.

6.1.5 Automatic Adaptation to User Interests by Observation

The user’s behavior, while exploring an exhibition, is constantly monitored (i.e. the time spend in front of different exhibits belonging to a certain category within the exhibition, the additional information requested by the user, aborted presentations) and analyzed by the system. Based on the data achieved in this way, the system finds an appropriate user group. In the prototype setup, the system monitors the user, until he reaches the end of the museum tour. Figure 6.8 below shows the a screenshot of the statistic analyzer module which illustrates the data collected by the server in the Völklingen Old Ironworks scenario.

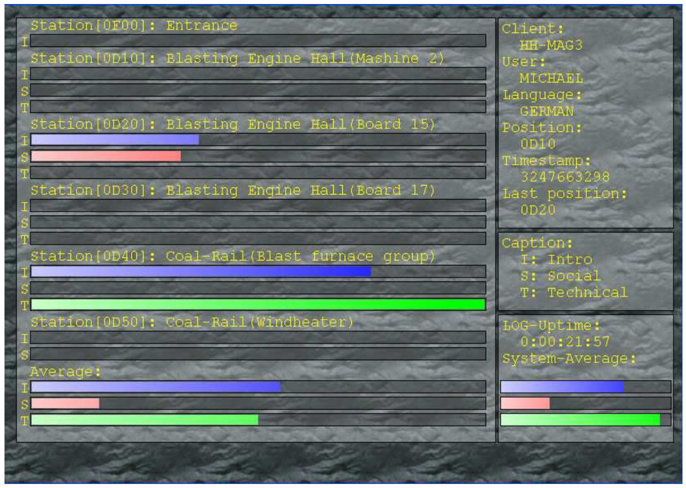


Figure 6.8: Screenshot of the statistic analyzer module showing user data

The statistic analyzer module categorizes users into different groups according to their interests shown while exploring the museum site. While the presentations on the mobile devices are not influenced by this categorization, the content to be presented at the stationary system is pre-selected by the system based on these calculated user interests. Whenever using a stationary system, the Migrating Character explains to the user that the content presented on the stationary system provides further details on exhibits and information which seem to be particularly interesting to the user. However, the users are always free to disregard this pre-selected content and make their own choices.

6.1.6 The Longtime User Model

In addition to the internal user model which is built up by the server during a user's visit at the museum, a second, external user model is utilized (see Kruppa, Heckmann, & Krüger, 2005). This user model serves as an initial setup whenever a user visits a museum. The actual user model is stored on an Internet server[9]. In this way, information gathered during previous museum visits may be used to come up with an initial categorization of a user in a previously unvisited museum. However, since the data collected may be sensible, users are always free to turn of the external user model feature. In this case, when starting a museum visit, the museum guide system will handle the user with a standard setup. In case the user agrees to share the information on the Internet, the initial setup sequence works as illustrated in Figure 6.9.

Users participating in this service are identified by a unique ID, the UID (User ID). Using a personal mobile device, the UID is stored on the device and automatically transmitted to the museum guide server. Otherwise, using a rented mobile device, the user is asked to enter the UID. Thereupon, the museum guide server requests available information about user preferences from the user model server. The acquired information is than integrated with the internal user

[9]The server is available online at: http://www.u2m.org

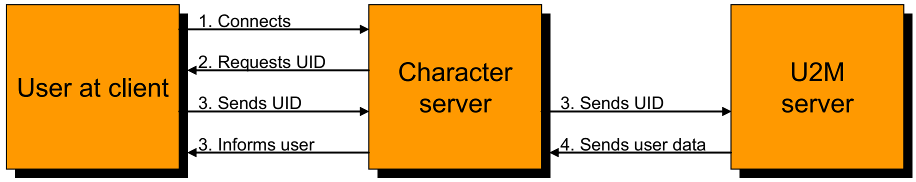


Figure 6.9: Initial user model setup sequence

model. Once this setup process completes, the user is informed about the initial setup chosen by the museum guide server. At this point, the user is free to change this setup or to completely disable the whole user adaptation process. A screen shot of an exemplary user model downloaded from the user model server and the corresponding dialogs on the mobile device are depicted in Figure 6.10.

Figure 6.10: User model representation at U2M.org and corresponding layout on a PDA

6.1.7 Template-Based Adaptive Video Documentaries

A major limitation of the mobile device used in the PEACH mobile museum guide is the small screen size. As mentioned in Sections 5.2, 5.1 and 6.1.2, the character layout had to be adapted in order to fit on the small screen of the mobile device. This was also necessary to leave sufficient screen space for visual information presentation. However, even the best layout cannot change the physical size and resolution of the display. Especially when presenting very detailed images with a high resolution on the mobile device's screen, the resulting quality may not be sufficient in order to show every detail of the image. To overcome this general limitation, a system for template-based adaptive video documentary generation (see Rocchi & Zancanaro, 2004 and Kruppa et al., 2003 for a detailed description of the system) has been adapted in order to work on a mobile device. Automatic composition of video documentaries is often done by selecting annotated, pre-recorded clips from a database which are concatenated to form a new movie. However, the approach of Rocchi and Zancanaro is based on a different idea. The building blocks of the generated video documentaries are shots which are defined beforehand by means of a scripting language. The shots themselves consist of camera movements which are applied to 2D images in conjunction with audio files. In the terminology of cinematography (see also Metz, 1974), a shot refers to a continuous view from a single camera without interruption. In the context of the video documentary generation system, a shot refers to a sequence of camera movements which are applied to the same image. Camera movements are available in three dimensions, corresponding to a camera's pan, zoom and tilt functionality. Between shots, different transition effects may be chosen:

- **Display:** The first frame of the new shot immediately replaces the last frame of the previous shot without any transition effect.
- **Cut:** Similar to the "display condition", however a white space is inserted between shots.
- **Fade:** An old shot is gradually replaced by a black screen (fade-out) or a black screen is gradually replaced by a new shot (fade-in).
- **Crossfade:** A new shot gradually replaces an old shot by combining a fade-out on the old shot and fade-in on the new shot.

A set of strategies similar to those used in documentaries are applied during the automatic composition of video documentaries. Two different classes of strategies can be identified within the system. The first class consists of constraints adapted from the grammar of cinematography. An exemplary constraint in this class restricts the use of a pan-left immediately followed by a pan-right camera movement. The second class consists of generally accepted conventions for camera movements in video documentaries. An exemplary rule from this class would encourage the use of sequential scene cuts when visually introducing different characters instead of using a fade effect.

The output of the automatic video documentary generation system is an XML-based presentation script (see appendix C for an exemplary video presentation script). In order to watch the generated presentations, an additional player component was developed. This player component interprets the presentation script and renders the corresponding video documentary. The communication between video documentary generator and player is realized via a standard TCP/IP socket connection. The player itself is designed as a component which can be integrated seamlessly into any GUI that is based on a technology allowing for Macromedia Flash MX content playback. The whole architecture of the automatic video generation and playback system is depicted in Figure 6.11.

As illustrated in Figure 6.11, the communication between the video documentary generation system and the video player is usually routed through a GUI in which the player is embedded. The XML script received by the GUI is simply forwarded via internal channels to the player.

In order to adapt the video generator system to work within the PEACH prototype, first the layout of the player component had to be changed in order to fit the small screen area available

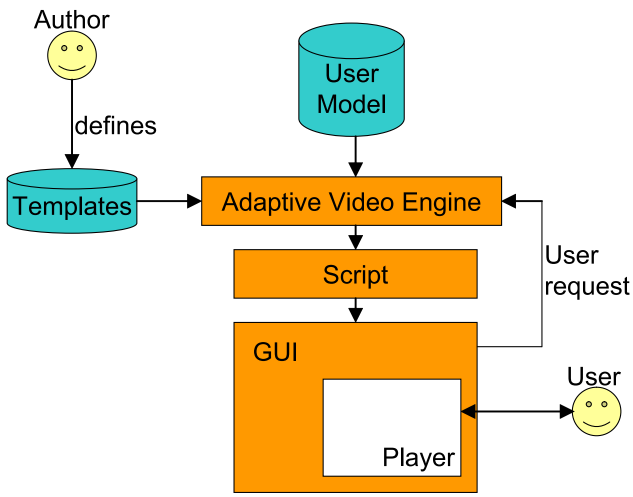


Figure 6.11: Architecture of the video documentary generation system (according to: Rocchi & Zancanaro, 2004)

on the mobile device. Since the video playback had to be integrated seamlessly into the whole presentation performed by the Virtual Characters, the communication mechanism between video player and video generator was also changed. Instead of a direct communication via the GUI, the PEACH server requests specific videos from the video generator. The generated video script is send back to the PEACH server which then incorporates the script in a surrounding presentation script for the mobile device (including instructions for the Migrating Characters and also slide show instructions). In this way it is possible to generate a coherent presentation, allowing the Migrating Character to announce a video clip, watch the clip together with the user and summarize the content as soon as the clip is finished (the integration of the video generator and player component in the PEACH prototype has also been described in detail in Rocchi, Stock, Zancanaro, Kruppa, & Krüger, 2004).

Figure 6.12: A migrating character announcing and watching a video clip

Figure 6.12 shows a sequence of screen shots of a character announcing a video clip and watching the clip together with the user. The implemented system allows to generate and integrate video clips into the PEACH museum guide which maximize the use of the mobile device's small screen by automatically showing relevant details of large scale images while playing back audio commentary.

6.1.8 Multi-user Support

A major problem in the PEACH scenario was the fact, that the number of Virtual Windows available in a museum will always be significantly lower than the number of visitors. This is due to both economical as well as practical reasons. It is hence desirable to find methods which will allow several users to use a stationary system at the same time. Within the scope of the PEACH project, different techniques have been developed to support multiple users interacting with a single, stationary device. When using standard devices like for example a touch screen in the realization of multi-user applications, the first problem is to find out which user is performing which action. There are specialized devices, like for example the MERL [10] Diamond Touch, which allows multiple users to interact with a single touch screen. However, interacting with a touch screen requires the users to stand directly in front of the screen, and hence they obscure part of the display for users standing behind them. In our scenario, we want to benefit of the fact that each user has its own mobile device. These devices may not only be used to present localized information throughout the museum, but may also serve as a user-interface when interacting with the stationary systems. In (Kruppa & Krüger, 2003), several different methods for a combined use of PDAs and large remote displays have been explored.

We identified two different techniques which do fit very well in our scenario of combined personal devices and public information systems.

For the first technique to support multi-user interactions, we adopt the metaphor of a remote control (this technique was described in detail in (Kruppa, 2004)).Users interact with the Virtual Window by pressing on buttons that are displayed on their mobile device. Using wireless LAN technology, this interaction is communicated via a server to the stationary system. This server also selects the content to be presented at the stationary system, based on the user interaction history. The user may choose different presentations, which are arranged in a list, sorted in order of highest interest.

When another user approaches the Virtual Window the presentation lists of all users in front of the Virtual Window are combined to form a new list, which holds only items which are of interest to all users. In case this list should be empty, very general presentations are included in the list, which should be of interest to most visitors of the museum. As soon as the running presentation is finished, the newly generated list is shown on each mobile device and on the Virtual Window (see Figure 6.13). At this point, each user chooses a presentation on the mobile device. After the first user has chosen a topic, a countdown is started on the Virtual Window. Each user may then make a decision until the countdown is finished or each user has made his/her choice. To ensure that there will be a definite decision, each user has a different weight in the voting process. The first user arriving has the highest weight, the last one the lowest. This also reflects the fact that the Virtual

[10] The Mitsubishi Electric Research Laboratory, http://merl.com

Window belonged to the first user, and hence this user still keeps the most control. However, this method does only work in cases where all the users sharing a single device have at least some overlapping interests in their user profiles. For completely heterogeneous user groups, sharing no interests at all, a different method is necessary. Instead of forcing users to agree on topics,

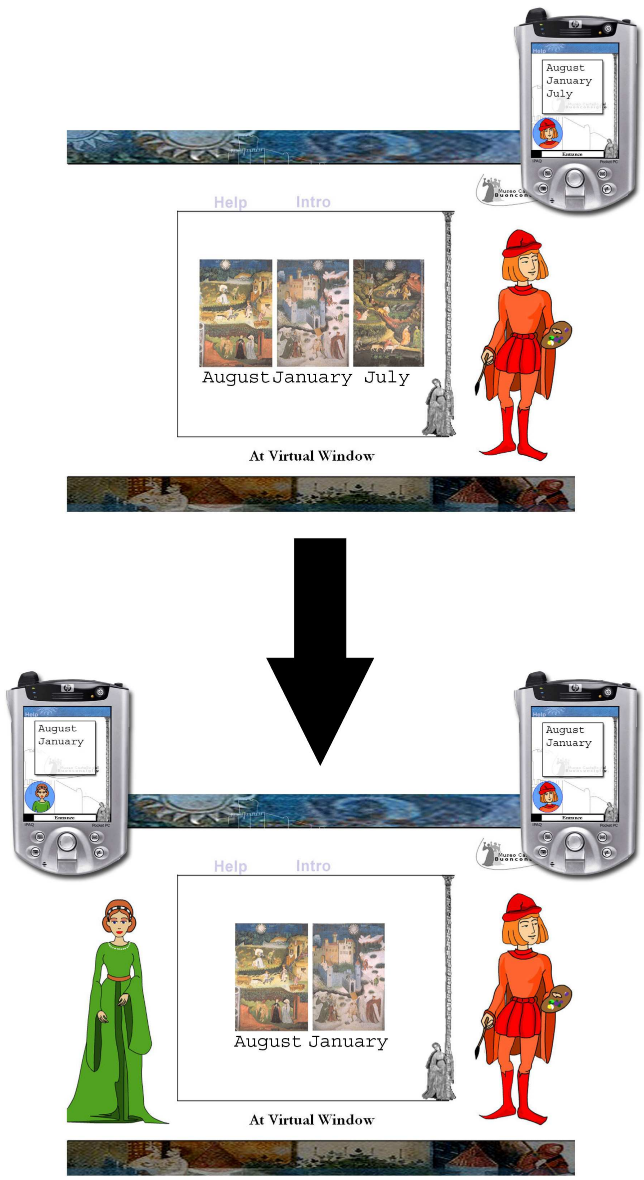


Figure 6.13: Multi-user interaction using a voting mechanism

we invented a system which combines personal devices and stationary presentation systems in order to provide presentations which fit the interests of all users sharing a stationary information system. The idea is to have a “root presentation” on the stationary presentation system and to fill in personal presentations on the user’s personal device, whenever the actual topic of the “root-presentation” does not fit the interests of the user (this approach has been discussed in detail in (Krüger, Kruppa, Müller, & Wasinger, 2002)). While the focus of the presentation on the stationary presentation system is on general information on a specific topic, presentations on the personal mobile devices present in-depth information related to topics mentioned in the “root-presentation”. The presentations to be shown to each user on their mobile devices are chosen automatically based on the user model derived during the museum visit. Users are free to stop a presentation on their personal device at any time, which will then result in an update of the user model of that particular user.

Since the “root-presentation” on the stationary device continues, while the individual presentations run on the users’ personal devices, this presentation method might possibly put a high cognitive load on the users. This is due to the fact that they have to concentrate on the presentation on their personal device while still hearing (and maybe also seeing) the presentation running on the stationary presentation system. However, empirical studies with this type of presentation system have shown that users are capable of concentrating on their individual presentations without being distracted by the “root-presentation”. The corresponding user study is discussed in Chapter 7 as well as in (Kruppa & Aslan, 2005). A critical point within these multi device presentations is the moment when users have to switch their attention from one device to another. The aforementioned user study has shown that the Migrating Character may successfully help users in guiding their attention from one device to another by migrating between the different presentation devices.

6.2 VRI - The Virtual Room Inhabitant

The Virtual Room Inhabitant (VRI) was developed within the scope of the Sonderforschungsbereich 378 at Saarland University. The main goal of the project was to realize a Migrating Character, capable of appearing and moving on arbitrary surfaces. The character was realized in the Saarland University Pervasive Instrumented Environment (SUPIE, see Section 6.2.2 and also Butz, Schneider, & Spassova, 2004). The room is used by many different project groups which have realized a number of different services in the in it. However, since some of these services (as well as some of the installed devices) are not necessarily visible at first sight, the idea of a VRI is to offer assistance to users which are unaware of these hidden possibilities. Whenever a user enters the SUPIE room, the VRI should be there to help the user fulfill a certain task or find a specific device or object. Figure 6.14 shows the Virtual Room Inhabitant next to a wall mounted LCD panel.

6.2.1 Realization

In order to realize our vision of a life-like character “living” in our instrumented environment, several software/hardware components were combined (see Figure 6.15). Each device has to be registered on our device manager as a service. The device manager, in combination with a presentation manager, grants access to all registered devices (details are described in Stahl, Baus, Krüger, & Schmitz, 2005). In this way, we are able to share our devices between several applications running simultaneously. The remote access mechanism is realized with Java Remote Method

Figure 6.14: The Virtual Room Inhabitant beside an LCD panel

Invocation[11] objects, which allow arbitrary applications to control remote devices as if they were locally connected.

We use two different technologies in order to determine the users' positions: Active Infrared beacons (IR beacons) and active Radio Frequency Identification tags (RFID tags). Most positioning systems equip the user with infrared batches or RFID tags and put respective sensors at the walls or ceilings in order to detect the presence and/or proximity of a user. In our approach the beacons and tags are mounted at the ceiling and the user carries a PDA with integrated sensors (the standard built-in infrared port of the PDA and a PCMCIA active RFID reader card plugged into the PDA).

A received (or sensed) tag or beacon indicates that a user is near the location of the respective sender[12] (in case of the Infrared Beacons, which need a direct line of sight, the orientation of the user/PDA is also determined). The active RFID tags have a memory of 64 bytes of which 56 bytes can be freely used (read and write). We use this memory to store the coordinate of the tag. Each IR beacon is combined with one RFID tag that provides the coordinates of the beacon and the tag itself. When a tag or beacon is sensed by the PDA a Geo Referenced Dynamic Bayesian Network (geoDBN, see Brandherm & Schwartz, 2005) is instantiated ("induced") and associated with the coordinate that is stored in its memory.

[11] http://java.sun.com/products/jdk/rmi/

[12] Or near a coordinate that lies in the area of the sender or receiver.

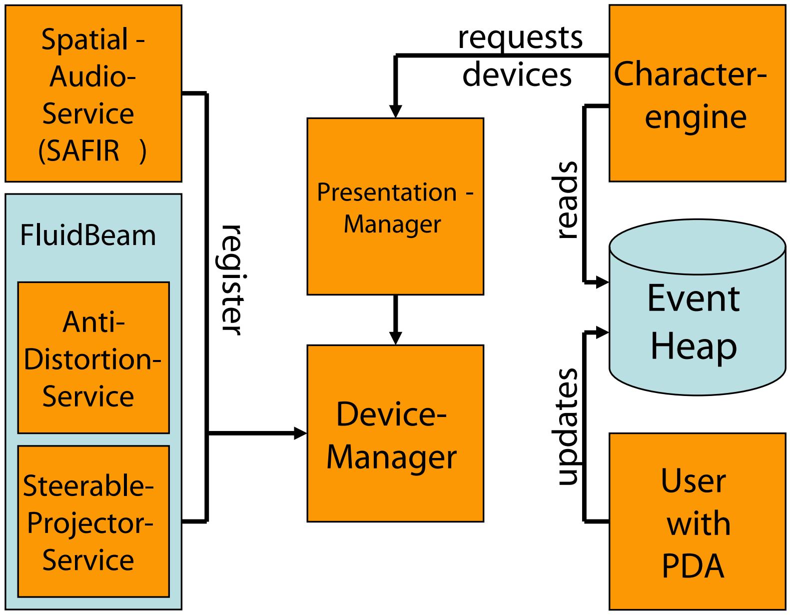


Figure 6.15: The system components of the Virtual Room Inhabitant

The position of the user is then calculated from the weighted combination of the coordinates at which geoDBNs exist:

$$\mathit{UserPos}_t = \sum_{i=1}^{n} \alpha\ w(\text{GeoDBN}[i])\ \text{Coord}(\text{GeoDBN}[i]).$$

n is the number of existing geoDBNs at time t ($n \geq \mathit{NumberOfReceivedSenders}_t$), (Coord(GeoDBN[$i$])) is the coordinate and w(GeoDBN[i]) the weight of the ith geoDBN. α is a normalization factor that ensures that the sum of all weights multiplied with α is one.

The calculated position is then forwarded by the PDA via wireless LAN to an event heap, a tuplespace infrastructure that provides convenient mechanisms for storing and retrieving collections of type-value fields (see Johanson & Fox, 2002). On this event heap, we collect all kinds of information retrieved within the environment (i.e. user positions, interactions with the system). Our central component, the character engine, monitors the event heap and automatically reacts according to changing user positions.

The VRI implementation combines a character engine with a spatial audio system and a steerable projector, to allow the character to freely move within the room (i.e. move along the walls of the room). These three main parts of our implementation will be explained in detail in the following subsections. However, we will start with a short description of the setup of our instrumented environment.

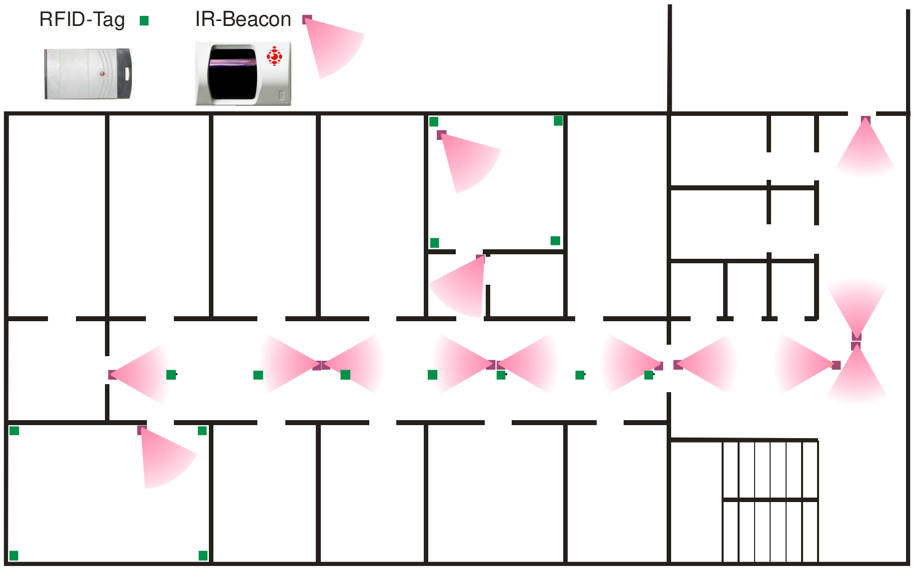


Figure 6.16: Exemplary infrared beacon and RFID tag positioning system setup

6.2.2 Exemplary Setup of an Instrumented Environment

The SUPIE room is located at the Artificial Intelligence Lab of Professor Wahlster at Saarland University. The main purpose of the room is to offer a hardware platform for developing and evaluating new technologies for instrumented environments. The SUPIE room is equipped with the following devices and sensors:

- a steerable projector combined with a digital camera
- another projector mounted on the ceiling projecting downwards on a desk
- a large scale, wall mounted LCD panel with a touch screen unit
- a spatial audio system
- a shelf, instrumented with an active Radio Frequency Identification (RFID) antenna
- a shopping trolley, instrumented with an active Radio Frequency Identification (RFID) antenna
- two steerable cameras
- Infrared beacons and active RFID tags for positioning purposes

6.2.3 Steerable Projector and Camera Unit (Fluid Beam)

A device consisting of an LCD projector and a digital camera placed in a movable unit is used to visualize the Migrating Character (see Figure 6.17). The movable unit (Moving Yoke) is mounted

Figure 6.17: Steerable projector hardware

on the ceiling of the instrumented environment and can be rotated horizontally (pan) and vertically (tilt). In this way it is possible to project at any wall and desk surface in the instrumented environment. The digital camera can provide high resolution images or a low resolution video stream which are used to recognize optical markers or simple gestures. As the projection surfaces are usually not perpendicular to the projector beam, the resulting image appears distorted. In order to correct this distortion we apply a method described in (Pinhanez, 2001). It is based on the fact that projection is a geometrical inversion of the process of taking a picture, given that the camera and the projector have the same optical parameters and the same position and orientation (see Figure 6.18). The implementation of this approach requires an exact 3D model of the instrumented environment, in which the projector is replaced by a virtual camera. Both the orientation and the field of view of the virtual camera must match those of the given projector. In this way we create a sort

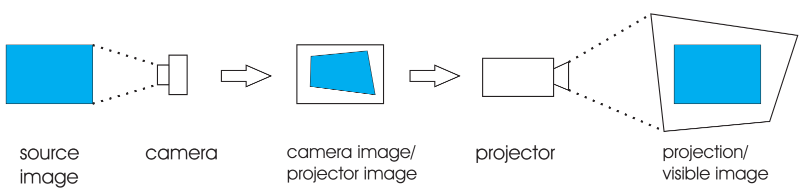


Figure 6.18: Distortion correction using the virtual camera method

of virtual layer covering the surfaces of the instrumented environment on which virtual displays can be placed. This layer is visible only in the area that is illuminated by the projector beam. Thus the steerable projector serves as a kind of virtual torch light making the virtual displays visible when it is moved in the appropriate direction. The displays are implemented as textured surfaces in the 3D model of the environment on which images, videos and live video streams can be displayed (for a detailed description of the Fluid Beam steerable projector and camera unit see also Spassova, 2004b and Spassova, 2004a).

The VRI is created as a live video stream texture on a virtual display. Thus it can be animated in real time by the character engine. By moving the virtual display in the 3D model and an appropriate movement of the steerable projector, the character appears to “move” along the walls of the room.

6.2.4 Spatial Audio

The synthesized speech of the VRI generates audio output that could be played by any device with audio rendering capabilities. In order to realize a Migrating Character capable of dislocating itself along arbitrary walls with a character voice coming always from the character’s actual location, we used the Spatial Audio Framework for Instrumented Rooms (SAFIR, (Schmitz, 2004)) that runs as a service in our environment and allows applications to concurrently spatialize arbitrary sounds in our lab. The character engine server (see Section 6.2.5) sends generated MP3 files, holding utterances produced by a speech synthesis system (the same system utilized in the PEACH project and described in Section 6.1.2.4) and the coordinates of the current location of the character to the spatial audio system, which positions the sounds accordingly. The anthropomorphic interface obviously appears more natural with the speech being perceived from the same direction as the projection is seen. This is particularly helpful in situations when other applications clutter up the acoustic space with additional audio sources at the same time. The spatial attributes of the audio output of the Migrating Character allow the user to associate the speech with the projection of the character more easily. Furthermore, it automatically directs the user’s attention to the position of the character when it appears outside the user’s field of vision.

The SAFIR system is a Java API for multi-channel spatial sound synthesis that permits to freely setup an arbitrary number of loudspeakers in a 2D (e.g. a ring) or 3D (e.g. a half-sphere) configuration, whereby a higher amount of speakers yields better results with respect to the spatial sound reproduction. The system is programmed in Java and is based on JSyn[13] libraries for the sound synthesis.

VBAP (Vector-Based Amplitude Panning, see Pulkki, 1997) is used as the core spatialization algorithm to position virtual sound sources. It provides computationally efficient functions to simulate the direction of a virtual sound source by selecting up to two (in a 2D setup) or three (3D setup) loudspeakers to emanate the signal with particular gain values to place the virtual sound between the active loudspeakers.

Since VBAP does not simulate distances of sound sources, the SAFIR system implements additional audio effects that create an impression of distance to the listener: The main distance cue is the decreasing intensity of sound with increasing distance due to air absorption, the system therefore attenuates virtual sound sources that are further away from the listener. Reflections of sounds that occur in rooms also provide cues that are interpreted by the brain to estimate the distance (among other attributes) of a sound source. To keep computational costs low, SAFIR only synthesizes the first reflection of a sound source, which comes from the same direction as the original sound source from the listener’s point of view. The system finally creates Doppler effects for moving sound sources which is perceived as a temporarily increased pitch when the source moves towards the listener or as a decreased pitch when it moves away respectively.

[13]http://www.softsynth.com/jsyn/

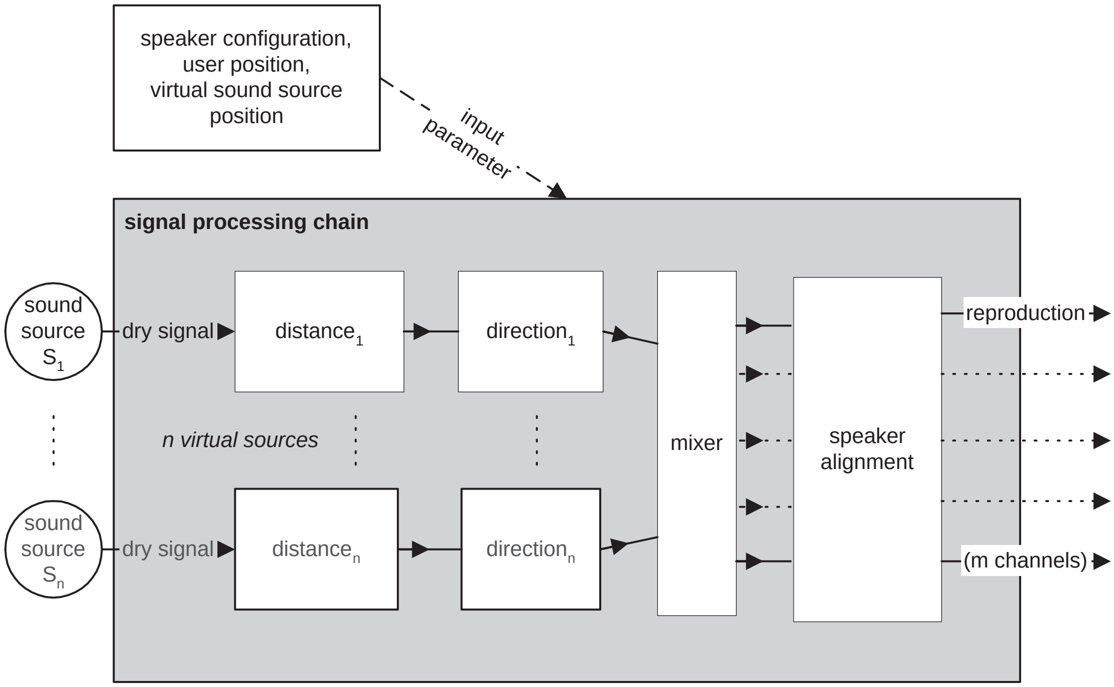


Figure 6.19: System structure of the Spatial Audio Framework for Instrumented Rooms (SAFIR)

The SAFIR system also utilizes the position of the user, if provided by the instrumented environment, by adjusting the speaker output (intensity and timing) according to their differing distances to the user. Information about the user position are also applied to the VBAP implementation, minimizing the direction error when the user moves away from the center of the room (the “sweet spot”).

Figure 6.19 visualizes the elements of SAFIR with the signal flow from system input (mono sound sources) to multi-channel output over loudspeakers.

6.2.5 Character Engine

The character engine consists of two parts, namely the character engine server (CE-server) written in Java and the character animation, which was realized with Macromedia Flash MX. These two components are connected via an XML-Socket-Connection. The CE-server controls the Flash animation by sending XML commands/scripts (see Appendix D for a complete example script). The Flash animation also uses the XML-Socket-Connection to send updates on the current state of the animation to the CE-server (i.e. whenever a part of an animation is started/finished). The character animation itself consists of ~9000 still images rendered with Discreet 3D Studio Max[14] which were transformed into Flash animations. To cope with the immense use of system memory, while running such a huge Flash animation, we divided the animation into 17 subparts. While the first consists of default and idle animations, the remaining sixteen are combinations of character

[14]http://www4.discreet.com/3dsmax/

gestures, like for example: shake, nod, look behind, etc. Each animation includes a lip movement loop, so that we are able to let the character talk in almost any position or while performing an arbitrary gesture. We have a top-level movie to control these movie parts. Initially, we load the default movie (i.e. when we start the character engine). Whenever we have a demand for a certain gesture (or a sequence of gestures), the CE-server sends the corresponding XML script to the top-level Flash movie which then sequentially loads the corresponding gesture movies.

The following is a short example of an XML script for the character engine:

```
<VRI-script>
    <script>
        <part>gesture=LookFrontal sound=welcome1.mp3</part>
        <part>gesture=Hips sound=welcome2.mp3</part>
        <part>gesture=swirl sound=swirlsound</part>
    </script>
    <script>
        <part>gesture=LookFrontal sound=cart.mp3</part>
        <part>gesture=PointDownLeft sound=cart2.mp3</part>
        <part>gesture=LookDownLeft sound=cart3.mp3</part>
        <part>gesture=swirl sound=swirlsound</part>
    </script>
    <script>
        <part>gesture=PointLeft sound=panel.mp3</part>
        <part>gesture=PointUpLeft sound=panel2.mp3</part>
        <part>gesture=swirl sound=swirlsound</part>
    </script>
</VRI-script>
```

Each script part is enclosed by a script tag. After a script part was successfully performed by the VRI animation, the CE-server initiates the next step (i.e. move the character to another physical location by moving the steerable projector and repositioning the voice of the character on the spatial audio system, or instruct the VRI animation to perform the next presentation part). As opposed to the approach in the PEACH project, the VRI does not use a ”real time speech synthesis” but instead it uses pre-synthesized speech which is was stored in an mp3 file. The PEACH approach could have been easily integrated here, however due to limited computing resources for the whole VRI setup, the pre-synthesis method was chosen instead. In order to guarantee for a smooth character animation, we defined certain frames in the default animation as possible exit points. On these frames, the character is in exactly the same position as on each initial frame of the gesture animations. Each gesture animation also has an exit frame. As soon as this frame is reached, we unload the gesture animation, to free the memory, and instead continue with the default movie or we load another gesture movie, depending on the active XML script.

In addition to its animation control function, the CE-server also requests appropriate devices from the presentation planner. Once access to these devices has been granted, the CE-server controls the spatial audio device, the steerable projector and the anti-distortion software. The two devices, together with the anti-distortion software are synchronized by commands generated by the CE-server, in order to allow the character to appear at any position along the walls of the instrumented room and to allow the origin of the character's voice to be exactly where the characters visual representation is.

Presentations within our instrumented environment are triggered by the user's movements within our lab. Our users are equipped with a PDA which is used for an outdoor/indoor navigation task. The indoor positional information retrieved by the PDA is forwarded to the event heap. On this heap, we store/publish all kinds of information which might be important for services running within the instrumented environment. As soon as a user enters the instrumented room, the CE-server recognizes the corresponding information on the event heap. In a next step, the CE-server requests access to the devices needed for the VRI. Given access to these devices is granted by the presentation manager (otherwise the server repeats the request after a specific period of time), the CE-server generates a virtual display on the anti-distortion software and starts a screen capture stream, capturing the character animation, which is then mapped on the virtual display (as described in detail in section 6.2.3). It also moves the steerable projector and the spatial audio source to a predefined initial position.

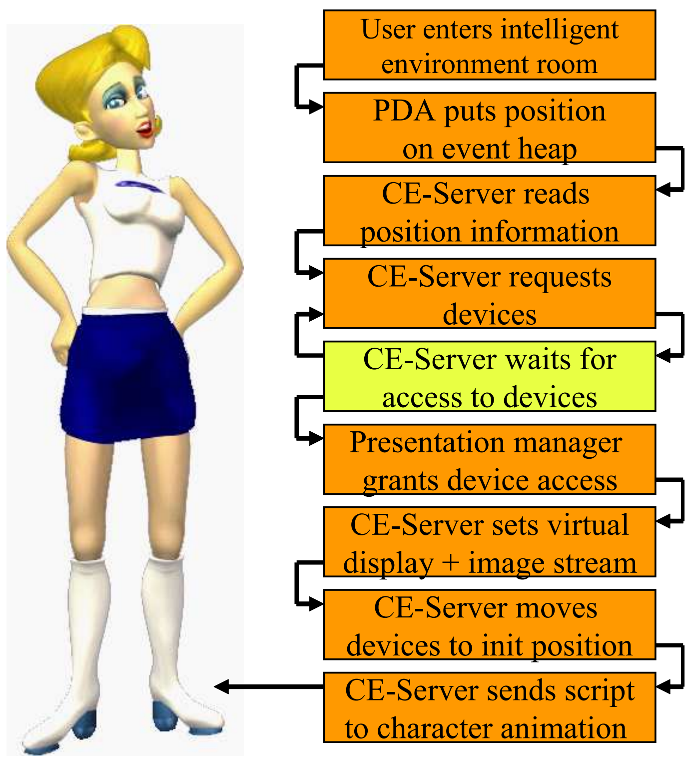


Figure 6.20: The steps taken by the system to initiate the Virtual Room Inhabitant

As a final step, the CE-server sends an XML script to the character animation, which will result in a combination of several gestures, performed by the character while playing synchronized MP3 files (synthesized speech) over the spatial audio device. The whole initialization process is indicated in Figure 6.20.

6.2.6 Character Movement

The VRI, when moving from one location to another is not walking, but we make use of a character animation which allows us to have a floating object instead. Prior to the actual character dislocation, the character is morphed into a very simple symbol, a circle (figure 6.21 shows the keyframes of the transformation animation).

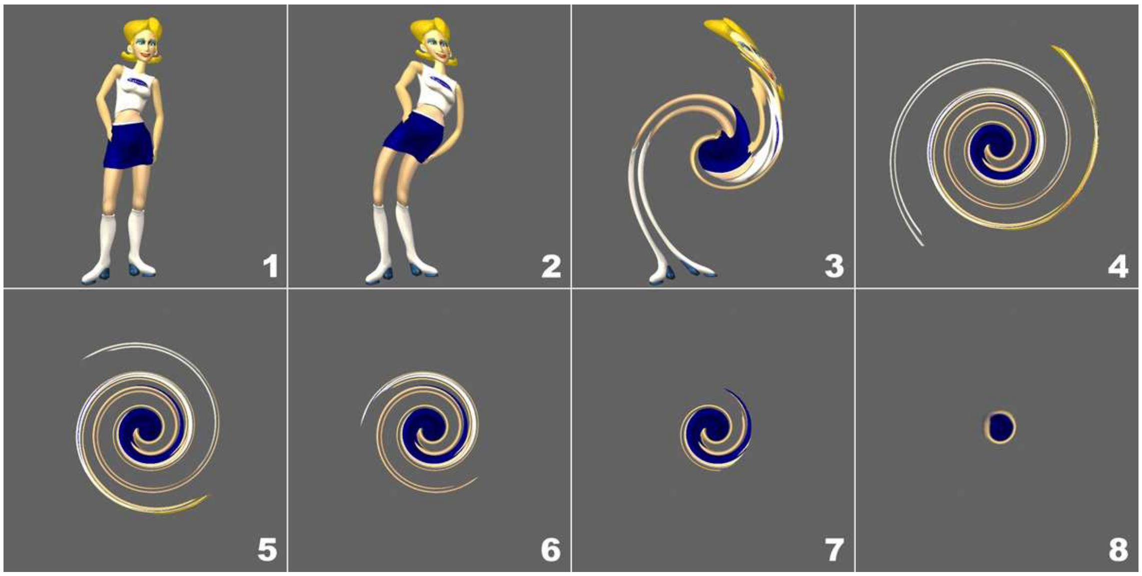


Figure 6.21: Keyframes of the character transformation prior to a character movement

This transformation is accompanied by a sound on the spatial audio system. The sound is located at the current position of the character projection. After the transformation from character to circle has finished, the projection of the circle is moved towards the target position. Again, a continuous sound on the spatial audio system moving along with the projection is used to give an additional, auditory clue to the user. When the character reaches its final destination, the initial transformation animation is played backwards, transforming the circle back to the character.

Using this method yields the following advantages:

- The character movement cannot be absolutely smooth, since the steerable projector movement isn't absolutely smooth either. Using a walk animation for the character movement would make this imperfect projection movement far more obvious.

- Using a walking animation for the character would limit the speed with which the projection could be moved, otherwise the character movement would appear unnatural. Using an animation where the character is capable of floating instead of walking allows to move the projection as fast as necessary.

6.2.7 Integration in an Airport Shopping Scenario

To test the VRI, we integrated it in a shopping and navigation demo running at our lab. In this scenario, users are given a PDA and perform a combined indoor/outdoor navigation task. The idea is to lead the users to an airport ground and upon entering the airport facilities to guide them towards certain duty free shops until the departure time is close. In the demonstration setup, these shops are represented by different rooms in our lab, one of them being the instrumented room. The VRI plays the role of a host, welcoming visitors and introducing them to the components of the instrumented room.

As soon as a user enters the room, the character appears on the wall nearby the entrance to welcome the user. Next step of the demonstration setup would be to interact with the smart shopping assistant (see Wasinger & Wahlster, 2006). The demo involves the instrumented shelf, recognizing if products are removed. This is realized by using passive RFID tags fixed to the products in the shelf. At the back of the shelf there is an active RFID antenna recognizing the products in the shelf. The antenna is sending information about the recognized RFID tags to a java server on a connected computer. When the system is started, the java server stores all the RFID tags received by the RFID antenna as an initial state. Later, when a product is removed from the shelf or put back into it, the java server generates events on the event heap (see section 6.2.1) which may be read by other applications in the instrumented environment. In addition, the shopping scenario also features an instrumented shopping cart. This cart features an active RFID antenna which works in the same way as the RFID antenna in the shelf and is also connected to the java server. An additional display mounted on the carts handle allows to display information on the products located in the shopping cart.

Involved in the demo is also a PDA which is used for navigation tasks as well as for the shopping scenario itself. The PDA is used to display information regarding the products in the shelf (see Figure 6.22 C). Furthermore, the PDA is also used for user interactions with the system. A flexible approach allows for several different modality combinations (which are discussed in detail in Wasinger, Krüger, & Jacobs, 2005). The system supports natural language input as well as so called intra-gestures and extra-gestures. While extra-gestures mean to physically interact with an object (e.g. picking it up from the shelf), intra-gestures relate to pointing gestures on a computer system (e.g. tapping on an object shown on the screen of the PDA). A user may for example pick up an object from the shelf and ask: “What is the prize of this camera?”. The audio signal is recorded and analyzed on the PDA and the derived information is combined with the information regarding the related gesture performed by the user. Also connected to the java server is a large LCD panel mounted on the wall right next to the shelf. The information regarding products is also shown on this display. When two objects are picked up at the same time, a product comparison is automatically shown on the LCD panel (and also on the PDA, as shown in Figure 6.22 D). Furthermore, the demonstration setup also includes so called Product Associated Displays (PAD, see Spassova, Wasinger, Baus, & Krüger, 2005). The PADs provide visual feedback to users interacting with the instrumented environment and in the special case of the shopping scenario they provide information while a user interacts with the shelf. When a product is taken out of the shelf, the generated event is picked up by the PAD server. A PAD is realized by utilizing the same steerable projector also used to realize the VRI. The PAD (e.g. a small projected display) appears right where the product used to be in the shelf and provides additional information on the product. Since the VRI and the PAD service both rely on the same hardware (i.e. the steerable projector), we make use of the mechanism described in section 6.2.1 which allows both services to share the same device. The steerable projector is also used to identify products in the shelf by putting a spotlight on them (see Figure 6.23). While the products have fixed locations in the shelf, the virtual representation of these products (i.e. photos or descriptions, see Figure 6.22 A and B) on the PDA can have varying orders (e.g. ordered by price, name or size). By tapping on the virtual representation of the product on the screen of the PDA and by saying: ”Find” in combination, the user may initialize this additional functionality. Since the user can not be expected to read

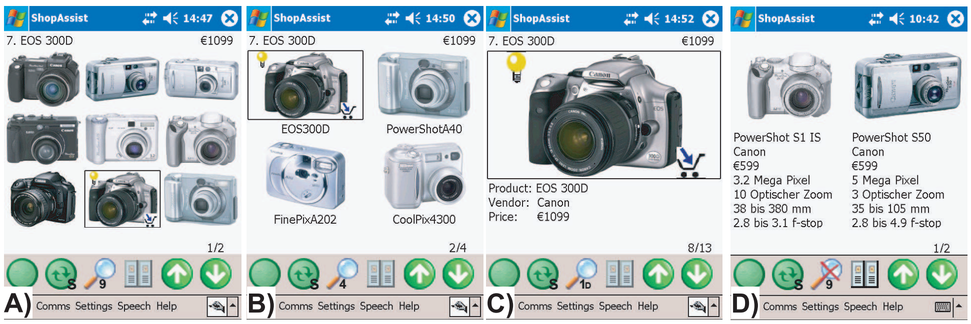


Figure 6.22: Screenshots of the product information shown on a PDA

such a long description of services and devices as the one above, the VRI is used to introduce the aforementioned devices and services step by step and describes the interaction possibilities given to the user. While explaining the demo, the VRI moves along the walls in order to appear next to the objects it is talking about. A user can stop the introduction by doing a wipe gesture across the VRI, who will then remain idle until it is reactivated by another wipe, which will let the VRI continue at the current step of the demo. Figure 6.23 shows the setup of the airport shopping scenario.

Figure 6.23: Left: the VRI assisting a customer, right: a particular object highlighted in a shopping shelf

To conclude the shopping demo, the VRI is triggered once more to notify users about the immediate boarding of their flight. It appears alongside the exit of the room, points to it and instructs the user to proceed to the boarding gate.

6.3 RefFind - Physical Object References with Migrating Characters

The RefFind project (see also Kruppa, 2005; Kruppa & Krüger, 2005) deals with the problem of performing references to physical objects in mobile application scenarios by utilizing different hardware setups in combination with Migrating Character technology. We have developed several different techniques which will allow a Migrating Character to perform a unique reference to a physical object. These different techniques, as well as a rule-based system determining the most suitable referencing solution in a specific situation, will be described in the following subsections.

6.3.1 Referencing Methods

The RefFind project builds on top of the technology developed in the PEACH project and the VRI project. We utilize the same basic setup as in PEACH, namely large, wall mounted displays and PDAs. We also use the same positioning technology based on infrared beacons. In addition, we added the hardware and software used and developed for the Virtual Room Inhabitant in order to allow a Migrating Character to appear on a wall by means of a projection and to allow the character to move along the wall by moving a steerable projector. The combination of the hardware and software technology of these two projects opened up new ways for performing unique references to physical objects.

Two referencing methods were directly derived from the PEACH prototype:

- Using the PDA, the Migrating Character can point towards a visual representation of the physical object. This visual representation, shown on the screen of the PDA, may either be an abstract graphic representing the physical object or a photograph of the actual physical object.

- In case there is a wall mounted display close to the physical object which is to be referenced, the Migrating Character can utilize the character transition technology in order to move closer to the target object. Once the character is located on the wall mounted display, it may point into the direction of the target object.

Additionally, we added the projection technology and the spatial audio system developed for the VRI to the RefFind project setup. In combination with the character device transition technology developed in the PEACH project, this allows a Migrating Character to disappear from the PDA of its user and reappear on the wall, were the target object is located. The character may then move as close to the target object as possible (this movement is limited by the projector technology as well as the physical setup of the room, since there may be objects in the projection way).

The simplest, yet sometimes sufficient method for performing a reference to a physical object in the RefFind project is to let the Migrating Character perform a sole utterance describing the object and its location.

6.3.2 System Setup

As mentioned above, the RefFind project combines techniques developed within the PEACH (see Section 6.1) and the VRI project (see Section 6.2) with a rule-based system to determine an optimal referencing strategy. In order to clarify the whole setup of the system, we will give a quick overview on the components and the underlying structure of the RefFind prototype in this section (Figure 6.24 depicts the components forming the RefFind prototype as well as the communication channels between those components). As a central component, the Migrating Character server developed for the PEACH museum guide is being deployed. The server is fed with a reference goal (this is only done for evaluation purposes while in a real life system, the reference goal would be determined automatically based on specific presentation goals of the Migrating Character). In addition, the Migrating Character server also receives sensory data regarding user locations. We make use of the dual sensor approach described in Section 6.2.1 to detect a user's position. Furthermore, the Migrating Character server is also retrieving information regarding the status of relevant objects in the environment. The Migrating Character presentations are realized by utilizing the device manager and character engine of the VRI project (see Section 6.2.5). In addition to these

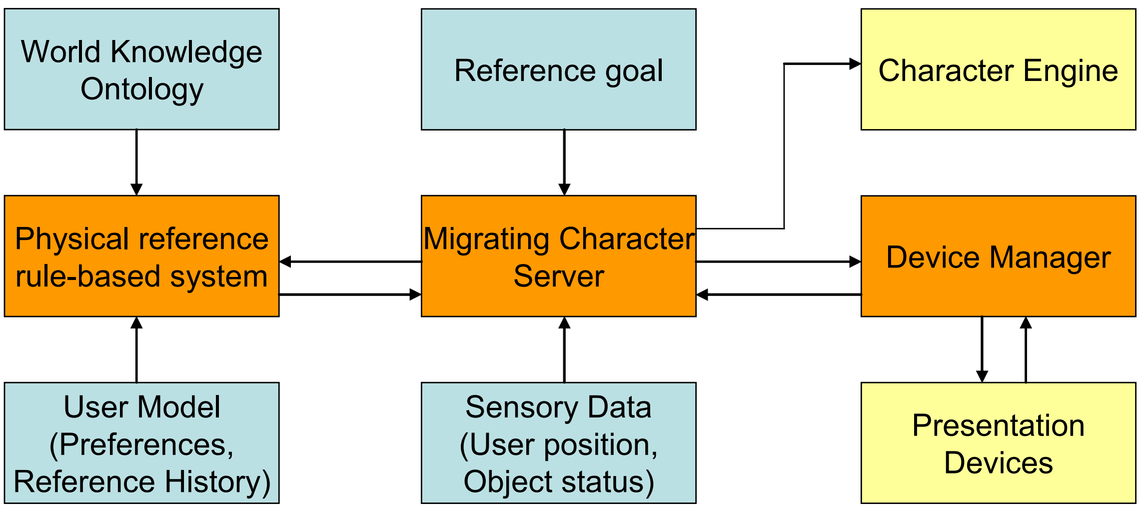


Figure 6.24: Overview on the different components of the RefFind system

previously developed components, a world knowledge ontology, a very simple user model and a physical reference rule-based system were added to the overall setup. In the following subsections, we will give a detailed overview on each of these newly developed components.

6.3.3 World Knowledge Ontology

We store the necessary world knowledge in an ontology which is accessible online[15] (see also Heckmann, Schwartz, Brandherm, Schmitz, & Wilamowitz-Moellendorff, 2005 and Heckmann, 2005) via the same server we also used for the long-term user models in the PEACH project. An

[15] http://www.u2m.org

online editor is used to set up the ontology and to put initial data into the ontology. Once the world knowledge is described, it may be accessed and modified via simple http requests. The result of such a request is an XML formatted document holding the desired information. At the moment, we have three different categories in the ontology, namely rooms, objects and users. Rooms and objects are defined by their size and location. In addition, objects have further information on their type, which is important when determining possible ambiguities while planning an object reference. In case objects are technical devices that can be used to display a Migrating Character, the ontology also allows to mark them as “in use” or “vacant”. Objects are always associated with rooms in the ontology and their positions are relative within the room coordinates. Information on users is stored in the ontology along with their preferences, location, orientation and history of references. In this way, the objects that are currently relevant at the user’s position may be easily determined. The ontology structure also allows to easily find out whether a user needs to dislocate herself in order to be in reach of a certain object.

6.3.4 Reference History

A history of the last two objects referenced is kept for each user. This history may help to simplify references to objects which have recently been in the focus of the user. In some cases, when referencing an object which is represented in this history, the Migrating Character may perform a unique spoken reference, which would have otherwise been ambiguous. If there are two objects of the same kind in the history, the spoken reference would have to be very precise, like for example: “The last but one image we have seen”. If the reference history holds only a single object, or two objects of different kinds, the spoken reference may be reduced, for example: “Back to the previous painting”.

6.3.5 A Rule-Based System To Determine An Optimal Reference Strategy

Whenever a physical object reference is necessary, a rule-based physical reference system is instantiated. The data fed into the system is taken from the online ontology and transformed into facts which are then asserted into the rule-based system and evaluated by the defined rules. These facts describe the room in which the user is located at the moment, the objects located in that room, the reference goal and the actual situative context of the user. Whenever instantiated, the rule-based system will determine the next step which is necessary in order to reach the final goal of a unique physical object reference. Based on the result produced by the rule-based system, the Migrating Character will perform the according actions and then (in case this wasn’t the final step of the reference task), the rule-based system will be instantiated again with the updated world model and user context. This process is illustrated in Figure 6.25.

Starting from the initial reference goal, the first three rules within the rule-based system determine whether a user dislocation is necessary (i.e. user must move to another room, turn towards another wall or move closer to a wall if the target object is too small from the user position). If any one of these three rules is activated, the result determined by the rule-based system will not be a reference instruction for the Migrating Character but instead it will be an instruction for a necessary physical user context change. Based on this instruction, the user is asked to relocate accordingly by the Migrating Character. Whenever the user context has changed, the rule-based

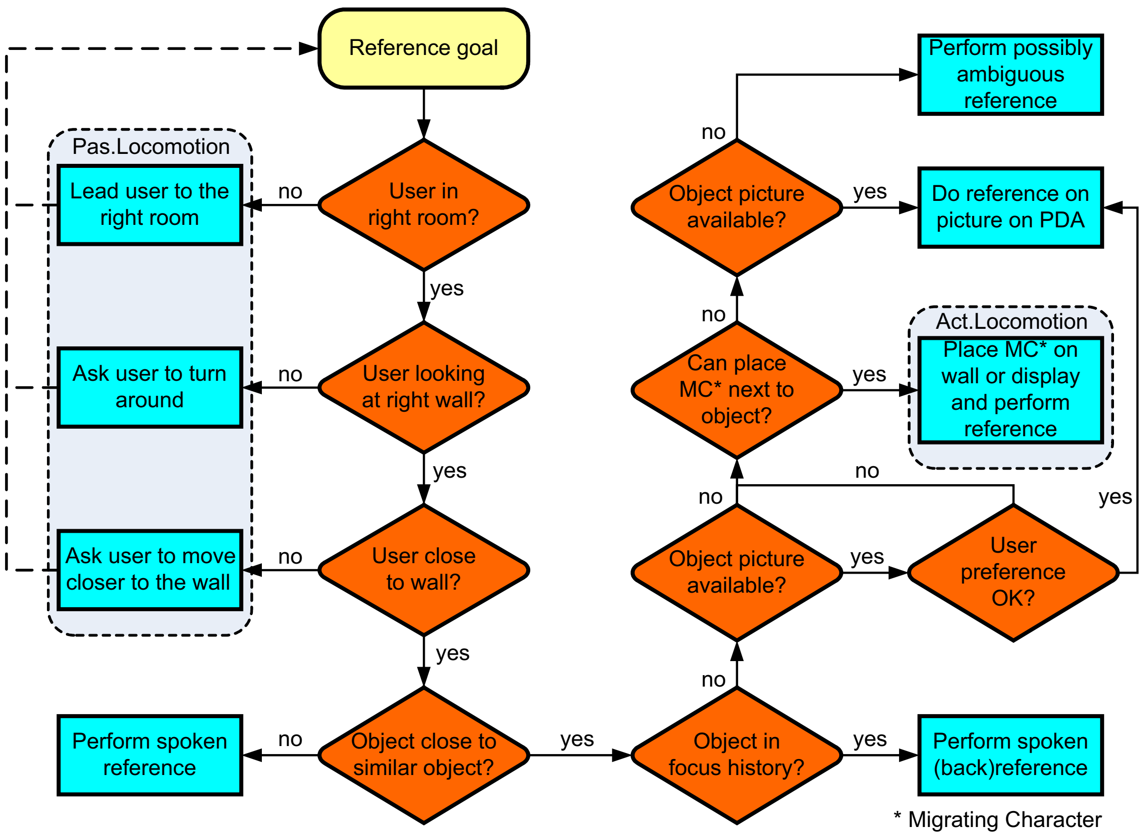


Figure 6.25: Schematic overview on the reference method determination process

system is restarted. Once, none of the first three rules are activated due to the user's physical context, the result determined by the rule-based system will be a reference instruction for the Migrating Character. In case the object to be referenced is not close to a similar object (this is calculated based on the size and type of the target object and surrounding objects), a simple spoken reference will be sufficient. Otherwise, based on the user's focus history (i.e. last two objects which have been referenced, see Section 6.3.4), the availability of a picture of the target object, the availability of the necessary hardware and physical space for a character dislocation right next to the object and the user's preferences, different strategies are chosen to disambiguate the physical object reference. If all strategies fail, the worst case is a possibly ambiguous spoken reference by the Migrating Character.

The rule-based physical reference system has been realized in JESS[16] and was integrated into a Java server which handles communication with the knowledge base and sensor services. Whenever necessary, the reference system is invoked by the java server. The reference system itself consists of about 20 rules and functions. The following exemplary rule was written in Jess and determines whether a user is looking at the right wall (i.e. the wall on which the target object is located). In case this is not true, a second rule (findTurn) is called which determines in which direction the user should turn in order to solve the problem:

[16] A rule engine for the java platform - http://herzberg.ca.sandia.gov/jess

```
(defrule userOrientation
    (phase idle)
    ?user <- (user   (name ?userName) (focus1 ?objectName)
                 (isInRoom ?room) (looksAtWall ?userWall)
                 (isAtStep detOri))
    (object (name ?objectName) (isInRoom ?room)
    (isOnWall ?objectWall))
    (room (name ?objectRoom) (numberOfWalls ?roomNumberOfWalls))
    =>
    (if (not (eq ?objectWall ?userWall)) then
        (printout t "User "?userName" is not looking at the wall
        where the object "crlf)
        (printout t ?objectName" is located. User must turn ")
        (bind ?temp (findTurn ?roomNumberOfWalls ?objectWall
        ?userWall))
        (if (= (abs(?temp) 1)) then
            (if (= ?temp -1) then (bind ?direction left) else
            (bind ?direction right))
            (printout t ?direction crlf)
            (modify ?user (isAtStep idle))
            (printout d "<decision user="?userName"
            action=mustTurn direction="?direction">")
        else
            (if (<= ?temp -1) then (bind ?direction right) else
            (bind ?direction left))
            (printout t ?direction crlf)
            (modify ?user (isAtStep idle))
            (printout d "<decision user="?userName"
            action=mustTurn direction="?direction">")
        )
    else
        (modify ?user (isAtStep detSize))
    )
)
```

6.4 Synopsis

The projects discussed in this chapter may serve as examples of the Migrating Character technology. They give an overview on what can be realized with these characters and they also provide ideas for technical realizations of Migrating Characters. However, none of the three projects does implement all the features included in the Migrating Character concept. While each of the projects focused on a different aspect of the Migrating Character concept, all together provide an almost complete coverage of the conceptual techniques discussed in Chapters 4 and 5. Table 6.1 gives an overview on the different Migrating Character concepts realized in each of the projects.

Even though the main motivation behind the development of the discussed application examples was the evaluation of the feasibility of the Migrating Character concept and not the development of a “Migrating Character Toolkit”, the described projects share a pool of fundamental technologies. Seen on a very abstract level, all described projects share the same basic setup, featuring a central server controlling the Migrating Characters behavior and also providing access to the characters knowledge-bases. All visual representations of Migrating Characters are realized as clients communicating with the central server via standard network protocols. While the developed characters vary to a large degree in both appearance and functionality, the underlying technology is still the same. All characters consist of separated animation files done in Macromedia Flash, usually one animation for each gesture. We use an approach similar to mobile agent technology which executes the agents code on different devices and for this purpose provides a runtime environment for the agent on each device. In the case of the Migrating Character examples discussed in this chapter, we provide a runtime environment by means of a Flash animation running on each client where a Migrating Character is to be executed. This Flash movie may load the Flash animations which constitute the Migrating Character at runtime whenever necessary. Furthermore, this Flash animation also handles the communication between the client and the server and also the communication between the user and the system (e.g. pressing on buttons). The central server in each of the described prototypes uses the same basic components for handling communication with the clients and the knowledge-bases. Other parts, for example the speech synthesis component in the PEACH project, the component which controls the steerable projector and spatial audio device in the VRI prototype, or the rule-based system determining optimal referencing solutions in the RefFind project were added to the core functionality of the server. The developed basic technologies in the described prototypes offer a solid basis for the implementation of Migrating Characters in general.

One aspect of special interest, namely multi-user interactions with public, stationary systems will be further investigated in the following chapter in which we present the results of a user study related to this topic.

		Realized in project		
		PEACH	VRI	RefFind
Migrating Character subconcept	Active character locomotion	⊗	⊗	⊗
	Passive character locomotion	⊗		⊗
	Migrating Character transformation for long distance movements		⊗	⊗
	Character transitions between devices	⊗		⊗
	Aural cues to support character transition	⊗		⊗
	Consistent character layout for mobile and stationary devices	⊗		⊗
	Consistent character voices on different devices	⊗		⊗
	Migrating Characters emerging from physical objects	⊗		
	Migrating Characters playing different roles	⊗		
	User guidance in Instrumented Environments	⊗	⊗	⊗
	Human-like physical object references			⊗
	Automatic choice of appropriate devices for a specific task			⊗
	Use of sensory data regarding the physical context	⊗	⊗	⊗
	Relaying of computing intensive tasks to stationary devices	⊗		⊗
	Adaptation to the user’s personal preferences	⊗	⊗	⊗
	User adaptation based on observation	⊗		

Table 6.1: Realized Migrating Character subconcepts in the different application examples

7 Experimental Evaluation

One of the main aspects of the Migrating Character technology is the integration of both mobile and stationary presentation devices into a single scenario. However, it is very likely that in most application scenarios, the number of users will probably be far bigger than the number of installed stationary presentation systems. When looking at the special domain of museums for example, we may notice that it should be possible for a museum to provide each visitor with a PDA to be used as a museum guide (as in the PEACH project, see Section 6.1) but for both economical and aesthetical reasons, installing the same amount of stationary systems within the exhibition space of the museum is simply impossible. It is hence desirable to allow not only a single user to interact with such installed stationary systems, but instead to provide some methodology which allows groups of users to benefit from a single stationary presentation system.

In Section 6.1.8 we described two different methods which allow for multi-user interactions with a single stationary system. Both methods are especially designed to be used in a scenario which combines mobile and stationary systems. While the first approach uses the mobile devices (each user in this scenario has his/her own mobile device) as some kind of remote control which allows the users to state a vote on upcoming presentations on the stationary system, the second approach uses both types of devices in a combined presentation mode. The main benefit of the second approach is that users do not have to agree on a single topic. Instead, they watch presentations which are rated to be of general interest, while their personal mobile device is showing additional presentations whenever a part of the running presentation on the stationary device is rated as uninteresting for a particular user. The Migrating Character of each user is used as a method to guide the attentional focus from one device to another during these multi-device presentations.

However, since the proposed presentation method has the potential of putting a high cognitive load on the users and thus might reduce the users' recall, we conducted an empirical user study into the effectiveness of such multi-device presentations. In the experiment we performed, we were especially interested in 1) the effect of the parallel presentation method on a recall task and 2) any effect that different methods for guiding the user's attentional focus in these complex presentations could have on the subjects. We believe the different methods for guiding the user's focus might possibly:

- Influence the cognitive load imposed on the user by the complex, combined parallel presentations;

- Have an effect on the time users need in order to switch their attention from one device to another;

Since life-like characters have proven to be very effective in guiding a user's attentional focus towards virtual objects in a virtual 3D world (see Lester et al., 2000; Towns et al., 1998), we expect the Migrating Characters to equally successful guide the user's attentional focus in our scenario.

Our belief that subjects should be capable of focusing their attention on the right device and not be distracted by the other is based on the fact that this work is closely related to the well investigated cocktail party effect (see Handel, 1989; Arons, 1992). Experiments have shown that human beings are capable of focusing on a single audio stream, while there may be many other streams (i.e. voices) at the same volume around (see Stifelmann, 1994). This capability is referred to as the cocktail party effect. In our experiment, we added a second information channel, namely a visual one, in order to find out whether subjects would be capable of focusing on a specific, combined audio/visual information stream among other audio/visual streams.

7.1 Method

Subjects participating in the experiment were organized in groups of three. Each subject was equipped with a PDA and they were positioned in front of a large LCD panel (see Figure 7.1). The experiment itself had three phases. In each phase, the subjects watched a movie clip on the LCD panel. After a varying period of time (as indicated in Figure 7.2), each subject was given one of two signals (a Migrating Character moving from the LCD panel to the PDA, or an animated Symbol on the PDA, or no signal at all as a control measure) telling subjects to focus their attention on the PDA from that moment on. On the PDA, the subjects had to follow a short presentation, while the presentation on the LCD panel was continuing. As soon as the presentations on the PDAs came to an end, subjects were signaled to focus their attention back on the LCD panel (i.e. the character moving back to the LCD Panel, an animated signal on the LCD panel, or no signal at all). Afterwards, subjects would continue to watch the still running movie on the LCD panel until it would finish.

All content presented on both types of devices was carefully chosen from a National Geographic Society publication (a DVD holding movie clips, as well as still images and written text) on explorers and discoverers. Among others, information about Robert Ballard, Richard Byrd and Louis Leakey was chosen to be presented to the subjects (see Subsection 7.1.2).

In order to figure out whether the combined presentation mode had an influence on the recall performance of subjects, a questionnaire had to be filled in after each presentation. The questions related both to content presented on the LCD panel (standard presentation mode) and on the PDAs (combined presentation mode).

In order to approximate the time subjects needed to switch their attentional focus from one device to another, we considered several possible ideas:

- Using an Eyetracker to find out where subjects are looking

Figure 7.1: Exemplary physical setup during the experiment

- Videotaping users and analyzing their gaze after the experiment was conducted.
- Showing memorable images on the PDA for a very short period, after subjects were signaled to focus their attention on the PDA

The first two options may be technically more precise. However, they only give a hint of the user's attentional focus, but they cannot tell whether a subject was just looking in a direction or actually perceiving information presented at that position. We opted for the third technique, after conducting a pre-test with seven subjects to find out whether the method worked at all. In order to be a little more precise, we decided to show each image for one second in a distracted form (i.e. upside down, stretched or inverted), and then for another second in the original state (figure 7.4 in Subsection 7.1.2 shows an example of the memorable images used). We included additional questions in the questionnaire to find out whether subjects had actually seen these images. Figure 7.2 illustrates the whole procedure of the experiment.

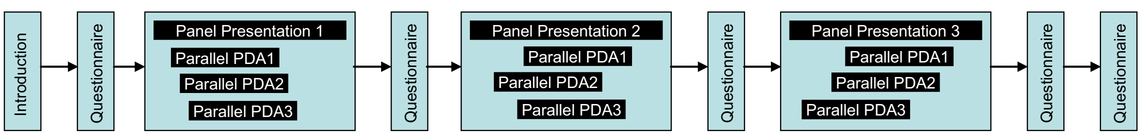


Figure 7.2: Procedure of the experiment

7.1.1 Subjects and Design

Subjects were 19 males and 23 females, all native speakers of German, recruited from the university campus and, on average, 28 years old. The subjects were paid 5 Euros for participation. The experiment lasted for approximately 40 minutes and was conducted in German. Subjects were organized in groups of three. The independent variables were Signal Method (character, symbol

Presentation part	Subj.1 signal	Subj.2 signal	Subj.3 signal
One: Ballard	Character	Symbol	None
Two: Carter/Byrd	None	Character	Symbol
Three: Leakey/Godall/Fossey	Symbol	None	Character

Table 7.1: Device switch signal method distribution throughout the experiment

or none) and Presentation Method (standard, running on the LCD panel or parallel, running simultaneously on the PDA and the LCD panel). We defined the subjects recall test performance on an assessment questionnaire and the estimated time for user focus shifts as dependent variables.

Both independent variables were manipulated within subjects. Table 7.1 illustrates how the variable Signal Method was manipulated during the experiment (i.e. each subject would get a different signal in each of the three phases of the experiment).

Each presentation part consisted of an initial presentation on the stationary information system, followed by a parallel presentation which was again followed by a roundup part on the stationary information system. The timing of the parallel presentation parts (i.e. the moment the parallel presentation starts) was slightly different for each subject, in order not to allow them to influence each other.

7.1.2 Materials

7.1.2.1 Presentations

The content presented to the subjects during the experiment was taken from a National Geographic Society publication, a DVD entitled: The Great Explorers and Discoverers. The content consisted of short video clips of approximately 4 Minutes length. Each movie clip focused on the work and life of up to three explorers or discoverers. We carefully tried to select content that presented information not commonly known in our culture area. In order to do so, we had pre-tests with 7 subjects. In these pre-tests, we tried to determine the amount of previous knowledge regarding the information delivered during presentations. Based on this data, we selected the following clips on explorers and discoverers:

- George Brass (who found a 3500 years old antique ship) and Robert Ballard (who found the wreck of the titanic)
- Howard Carter (who found the burial chamber of Tutankhamun) and Richard Byrd (who was the first man to fly to the South Pole)
- Louis Leakey (famous for his work on the human origin) and Jane Goodall (who became famous for her chimpanzee studies) and Dian Fossey (who studied the behavior of wild gorillas)

In addition to these video clips, the DVD also featured in depth information on each explorer and discoverer in the form of spoken text and photographs. We took this material to generate three different presentations with additional information on Robert Ballard, Richard Byrd and Louis Leakey. These additional presentations were shown to the subjects in the “parallel mode” on the PDA, while the main presentation on the stationary information system was going on.

7.1.2.2 Tests

Subjects were tested on their recall of the information presented with our parallel presentation system. The questions related both to visual (e.g. the question showed a photo of a person, object or an event and asked subjects to describe what they see) and audio information (for example dates and names mentioned during presentations). Figure 7.3 shows two typical questions from the questionnaire (original questions were in German).

1. Who is this and what is he doing?

Answer:

2. What did Byrd plan to leave at the South Pole?

Answer:	Learned from presentation	Pre-Knowledge	Guessed

Figure 7.3: Two exemplary questions from the questionnaire

Both types of questions were open-ended. For questions without visual input (i.e. images), we also asked subjects to indicate, whether they had previous knowledge on the subject, they guessed the answer or they actually remembered it from the presentation. Subjects were presented with an initial questionnaire asking for demographic information as well as for previous experiences with PDAs. The questionnaire also featured a short introduction to the first presentation part, indicating for each user how the presentations would be run and how they would be warned prior to a device switch during presentations.

After each presentation, subjects were given a questionnaire regarding the information delivered during that particular presentation.

At the end of the third presentation, subjects were given a second questionnaire regarding the effectiveness of each signaling method. The questionnaire included both objective and subjective questions. The objective ones asked subjects, whether they had seen the memorable images shown

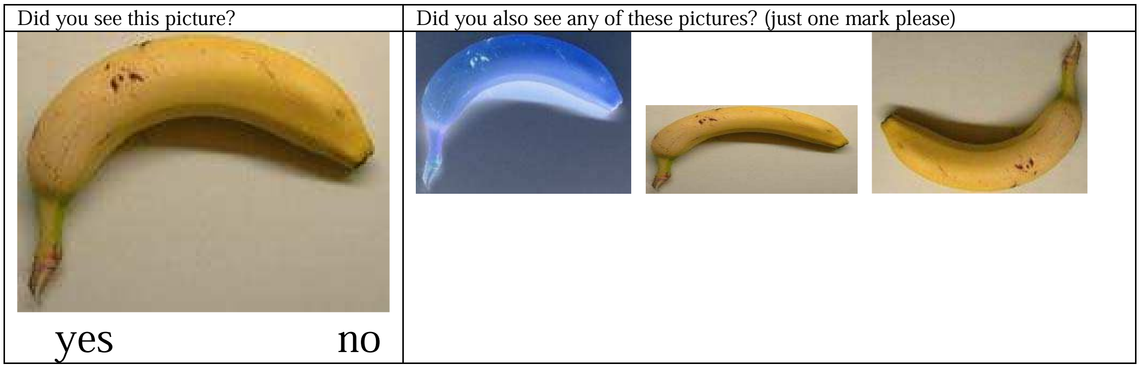


Figure 7.4: Example question to determine the time needed by subjects to switch their attentional focus

on the PDA at the beginning of each parallel presentation (see Figure 7.4). Subjective questions asked for a signaling preference as well as for general comments regarding the whole presentation system. The complete questionnaire is provided in Appendix A.

7.1.2.3 Apparatus

The hardware setup for the experiment consisted of a large, wall-mounted LCD panel, a spatial audio system and 3 Hewlett Packard iPAQs with integrated wireless LAN capability. A standard Windows PC was used to run the presentations on the LCD panel and to render the sound to the spatial audio system. The presentations for both mobile and the stationary devices were realized with Macromedia Flash MX. Each Flash Movie connected via a socket connection to a server implemented in Java. The server was run on a separate Windows machine and controlled the timing of the whole experiment. The server allowed for simple, text-based control by the experimenter, in order to start each presentation part after questionnaires were completed.

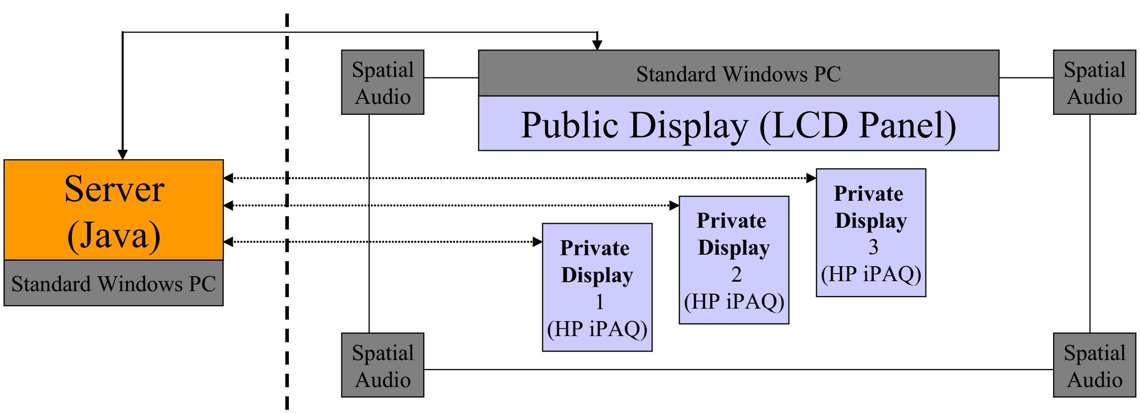


Figure 7.5: Apparatus Overview

In Figure 7.5 we present an overview of the apparatus setup. Important is the role of the server that is responsible for the synchronization of the single presentations. In this context the bi-directional connection between flash clients and Java server is significant, because in this way the flash clients are able to indicate whenever a presentation is about to end.

7.1.2.4 Operationalization of the Independent Variables

The Signal Method did not influence the way the content was presented to the users, however, the character signal and the animated symbol worked in a different way: the character disappeared on the "active device" (i.e. the one the user focuses on at the moment) and reappeared on the "target device" (i.e. the device, the user should focus his attention on next). The symbol animation on the other hand simply occurred on the "target device", so users had to constantly monitor the "passive device" in order not to miss the signal. Figure 7.6 illustrates both signal methods (on the left hand side, the animation steps of the character disappearing on the PDA and reappearing on the LCD panel are depicted, while on the right hand side, the key frames of the animated symbol for both the PDA and the LCD panel are shown). The second independent variable Presentation Method was manipulated during each presentation.

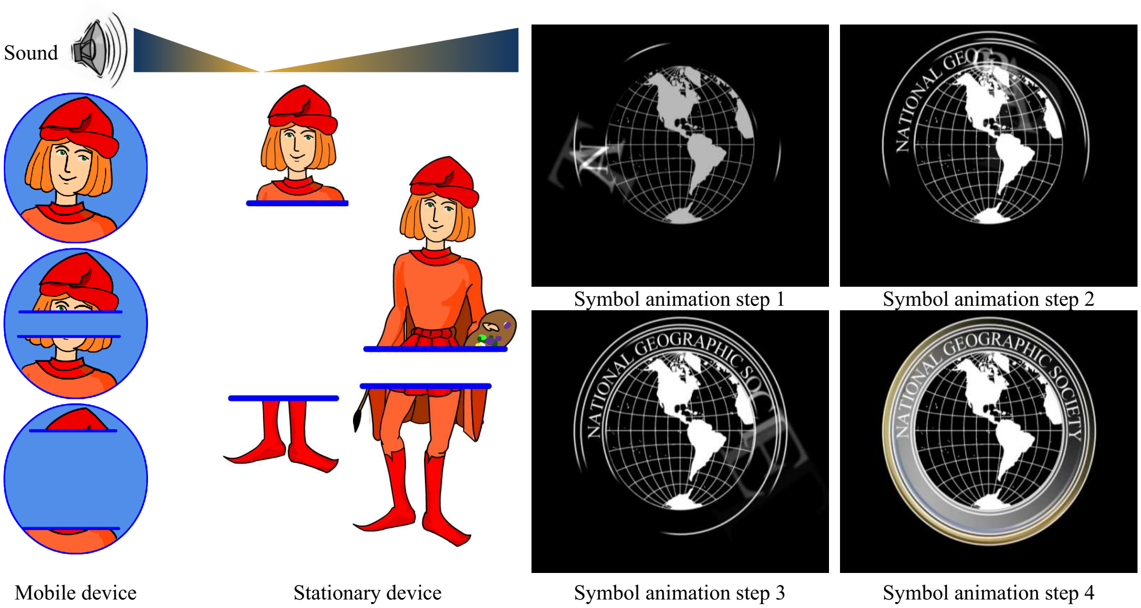


Figure 7.6: The different signals to indicate an upcoming device switch

7.1.3 Procedure

The experiment was run in the instrumented environment at the AI lab of Professor Wahlster at the Saarland University, Germany. During each experimental session, an experimenter was present in order to answer any possible questions and to control the flow of the whole session.

The subjects were told that the goal of the experiment was to evaluate the parallel presentation method. They were informed about the technical setup and about the way information would be presented to them. They were told that all information presented (i.e. images as well as spoken information) should be memorized as well as possible. In addition, they were told that after each of the presentations they would be tested on their knowledge regarding the information delivered during the precedent presentation. At the end of each questionnaire there was a short instruction regarding the next presentation part (these instructions informed subjects on the signal method that was assigned for them in the next presentation).

Just before the experimenter started the first presentation, he/she positioned the subjects in front of the LCD panel, gave a PDA to each of them (together with a "one-ear-headphone" which was connected to the PDA) and reminded them to always focus their attention on the right device (i.e. the "active device"). The experimenter also informed the subjects that the questionnaire might include questions regarding information they cannot know, since they were focusing on another device while that information was presented.

After each presentation was completed, subjects sat down at separate tables and answered a questionnaire with approximately 20 questions regarding the presentation just seen. Subjects could take as long as necessary to complete the test. After all subjects finished the test, the experimenter would ask whether they all understood the instructions for the next presentation and then he would position the subjects in the same order as before and equip them with the PDAs and headphones.

The experiment was concluded as soon as all subjects finished the additional questionnaire regarding the memorable images.

7.2 Results

In the following analysis an α level of .05 is used.[1]

7.2.1 The Visually Enhanced Cocktail Party Effect

The answers to the tests regarding the information delivered during presentations were scored by the experimenter. Each completely correct answer was awarded two points while partially correct answers were awarded one point.

For each subject, separate scores were calculated: three scores (one for each presentation / experiment phase) for the average performance in answering questions related to information presented on the PDA (i.e. parallel presentation mode) and another three for the average performance in answering questions related to information presented on the LCD panel (i.e. standard presentation mode). To evaluate the performance during parallel presentations, we subtracted the average means for the standard presentations from those for the parallel presentations. Figure 7.7 shows the average results of this calculation for each presentation. The graph indicates that there was a strong improvement in the performance related to parallel presentations from the first to the last presentation.

[1]Which means that if we observe an effect with $p < .05$, we may conclude that the probability that the sample means would have occurred by chance if the populations means are equal is less than .05.

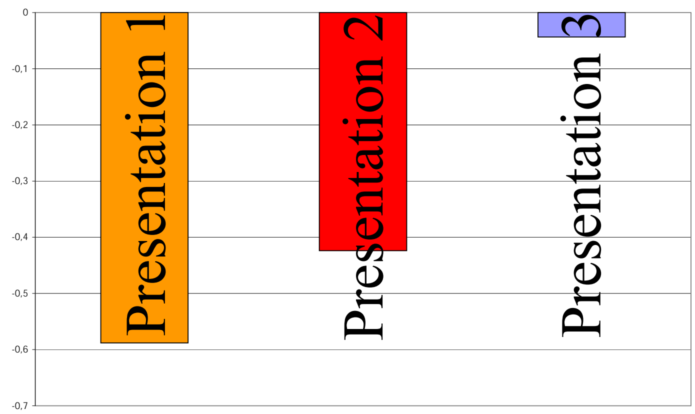


Figure 7.7: Difference between recall performance related to PDA and panel presentations

These data were than subjected to a paired samples t-test. The t-test testing the difference between presentation 1 and presentation 2 just very slightly missed the significance level ($t(42) = -1.757; p = .086$), however the t-test on the difference between presentation 2 and presentation 3 showed a highly significant result ($t(42) = -4.168; p = .000$). Also, the t-test testing the difference between presentation 1 and presentation 3 showed a highly significant result ($t(42) = -7.581; p = .000$).

Thus, the analysis showed a positive learning effect among subjects during the experiment regarding the parallel presentations. After only two presentation runs with the new presentation method, subjects learned to efficiently focus their attention on the device which was presenting relevant information to them. In the third presentation run, the subjects' recall performance regarding information delivered during parallel presentations nearly reached the performance level during standard presentations.

In order to find out whether the user focus guidance worked throughout the experiment in general, two different scores were calculated for each subject:

- The overall recall performance related to information delivered on the device users should focus their attention on
- The overall recall performance related to irrelevant information (i.e. information presented on the device users should not focus their attention on)

A two-tailed t-test comparing the two values revealed a highly significant difference between the two values ($t(42) = -16.902; p = .000$). From this result we conclude that the subjects' acceptance of the system's active attentional focus guidance was very high.

7.2.2 User Focus Guidance Methods

In order to analyze the impact of Signal Method on the recall performance of subjects during parallel presentations, average scores for each standard presentation were calculated for each subject and subtracted from the average scores corresponding parallel presentations (in this way we make sure that the result is not influenced by the complexity of the different material presented). These scores were than ordered according to the signal method used while achieving the score. Figure 7.8 shows the result of this calculation.

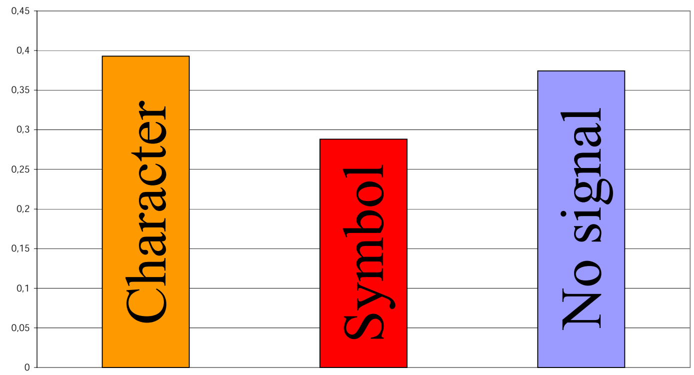


Figure 7.8: Recall comparison between different signals during parallel presentation

The graph indicates that subjects performed best when the Character was used to guide their attention. The performance during parallel presentations with no signal method was only slightly worse. However, the performance during parallel presentations with the animated Symbol was obviously inferior compared to the other two methods. The data was subjected to paired samples t-test. However, none of the t-tests comparing the different signal methods revealed any significance:

- Character vs. Symbol: $t(42) = 1.150$;$p = .257$
- Character vs. no Signal: $t(42) = .193$;$p = .848$
- Symbol vs. no Signal: $t(42) = -.933$;$p = .356$

Even though the t-tests did not reveal any significance of these results, we believe that we can base the following assumption on the data: When using a signal method which forces the subjects to split their attention between devices (as with the Symbol, which obliged subjects to monitor the “inactive” device in order not to miss the signal for an upcoming device switch), the subjects overall recall performance may be decreased.

In order to determine the most effective signal method with respect to the time subjects needed to move their attention from one device to another, we calculated average means for the memorable images shown prior to the parallel presentations on the PDAs. The results in the additional questionnaire were assigned different values: 2 points, where the subject recognized the distorted image (which was harder, since it was shown at the first second after the signal occurred), 1 point, where the subject recognized the original image (shown for one second, after the distorted image had been shown) and 3 points, if both images were recognized.

However, the results show that this task did not work (probably, because the times chosen for the memorable images were too short in general) for most subjects. The average score among all subjects for this task is only 0.75. Even more, there was not a single subject scoring two or three points more than once during the experiment. This data, from our point of view does not reveal any insight, apart from the fact, that the period of two seconds seems to be too short for most subjects to move their attention. Results for the three methods slightly varied (Character: 0.74; Symbol: 0.66; no Signal: 0.81) but this is not relevant in the light of the overall results regarding this task.

7.2.3 Subjective Assessment

The final questionnaire allowed subjects to state their preference for either one of the three signaling methods. They were also asked to give a reason for their decision. In addition, subjects were asked whether they had any criticism regarding the experienced presentation method. Figure 7.9 shows the subjective preference of the subjects regarding the signaling method.

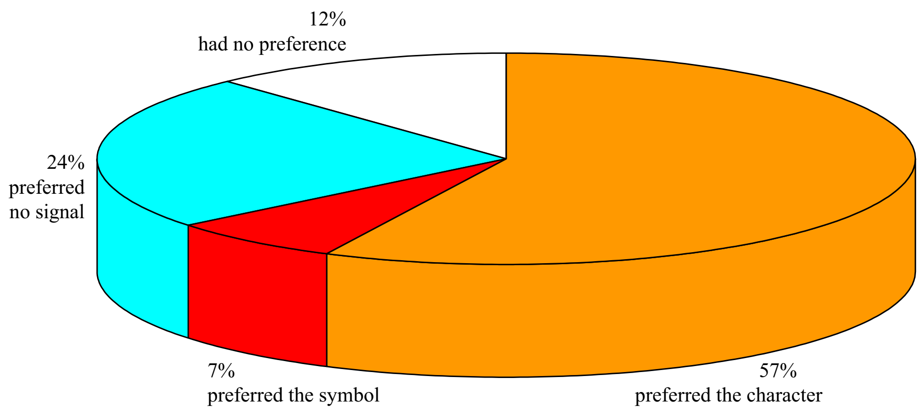


Figure 7.9: Subjective means regarding signaling methods

Common reasons given for the character preference were:

- The character is the most comfortable signal method, since it allows users to focus all their attention on just one device.
- The character is very eye-catching, as soon as it starts to appear/disappear.

Those subjects who preferred the "no signal condition" often complained about being distracted by either one of the visual signals. The only reason given for preferring the animated symbol was based on the fact that the symbol was nicely integrated in the overall layout of the application.

The general criticism stated by only 18 subjects revealed no insights:

- Five subjects complained that the sound for the panel presentations was too loud while two subjects mentioned that the sound was not loud enough (this had no influence on the subject's performance during parallel presentations, however those who said the sound was too loud, performed below average during panel presentations, while those who said it was not loud enough, performed exceptionally well).

- Three subjects felt, that there was too much information in the presentations.
- Eight subjects were distracted by either the background noises, the signals presented to the other users or by the sound of the panel presentation while trying to follow the presentations on the PDA.

7.3 Synopsis

The data supports our hypothesis that subjects would be capable of focusing on a single multimedia stream within a number of multimedia streams. The results clearly indicate that subjects were capable of concentrating on one device (during parallel presentations) and ignoring the presentation on the second device. However, we noticed a significant improvement in the subjects' recall performance with respect to parallel presentations during the experiment.

Based on these observations, we believe that in order to build public information systems capable of supporting heterogeneous user groups, our approach presented in this paper is very promising. Due to the strong learning effect regarding parallel presentations, we would suggest supporting users with a “training phase” prior to presenting critical content. In this way users could get used to the new presentation method without missing important information.

We believe that the weak results we saw regarding the time subjects needed to move their attention from one device to another was basically due to a design error in the experiment. Even though the pre-test subjects performed pretty well on this task, for the majority of subjects one or two seconds is obviously a too short period of time to focus their attention on a new device. The data did not support our hypothesis that different signaling methods to indicate an upcoming shift from one device to another would actually influence the time subjects needed to focus on the new device. As a design criterion for parallel presentations based on our experimental data, a minimum delay of three seconds is necessary when switching the presentation from one device to another.

The subjective assessment shows a clear preference for the virtual character as signaling method. Opinions stated by subjects in the questionnaire indicate that the virtual character was putting less stress on the subjects than the other methods.

The data showed no statistically significant impact of the variable Signal Method on the recall performance of subjects during parallel presentations. However results indicated again that the animated Symbol showed the weakest performance. The performance of the character condition and the one without a signal were almost identical. Nevertheless, when considering this together with the fact that the majority of subjects stated in the final questionnaire that they preferred the character as the signal method, we may conclude that a virtual character may help users to follow complex, multi-device presentations without putting additional cognitive load on the users. However, since quite a few subjects also preferred the condition with no signal at all, we think both options should be available for users to choose from.

Although generalization of these results should only be done very carefully, they indicate some important implications for software development aiming at supporting heterogeneous user groups sharing a single public presentation system:

- By combining a public multimedia information system with private PDAs, it is possible to generate presentations which should support all the different interests of the users sharing the public device.

- Parallel multimedia presentations offer, after an initial training-phase, the same potential to deliver information as a standard, single device multimedia presentation.

- In order to allow users of the system to follow multi-device presentations without putting too much stress on them, an appropriate way of guiding the users attentional focus needs to be used. A virtual character migrating between the devices promises a high potential to fulfill this task, even though there may be users who feel distracted by visual signals and hence prefer to do without any visual indication.

- When designing multi-device presentations, one has to keep in mind that there is a certain amount of time needed by average users to move their attentional focus from one device to another. A minimum delay of three seconds prior to presenting relevant content is necessary.

8 Conclusions

In the work at hand we developed the Migrating Character concept which realizes life-like character technology for application scenarios in which both mobile and stationary devices are used in conjunction. The resulting technology allows for the implementation of life-like characters, capable of assisting and guiding users while exploring physical environments. By bringing together the research areas of life-like characters and mobile computing, limitations of both areas had to be addressed, as well as new problems arising due to the combination. On the other hand, there are many beneficial aspects in favor of this new technology combination which we identified and discussed throughout this thesis.

On a practical level, the Migrating Character concept as well as the implemented prototypes allow for the rapid development of life-like characters for the domain of mobile computing and especially for the use in instrumented environments. The presented prototypic Migrating Character implementations give an impression of what may be realized with this technology and they also give an indication of the problems which may occur during the development of Migrating Characters.

In the following, we will summarize the most relevant results achieved throughout this work in detail. We will discuss the problems addressed as well as the solutions provided. We will conclude the chapter by discussing opportunities for further research based on the findings presented in this work.

8.1 Scientific Contributions

The thesis has introduced the new concept of Migrating Characters. We have motivated the idea behind the Migrating Characters in the introduction, we have introduced the Migrating Character concept and we presented new technologies for the realization of Migrating Characters which were analyzed in various case studies and we presented the results of an empirical user study focusing on one particular aspect of Migrating Characters. Specifically the following results are worth highlighting:

- **An analysis of the different aspects that constitute a life-like character implementation**

  In Chapter 2, we presented an analysis of what factors and aspects form the underlying structure of a life-like character. We discussed the different categories of life-like character implementations and the corresponding terms for those different character technologies as introduced by the corresponding researchers. We reviewed different opinions of researchers trying to determine the factors which define the believability of life-like characters. Apart from the characters' behavior, trying to imitate human behavior, the use of natural language in combination with fitting gestures, facial expressions and body language, were identified as important aspects towards increased believability of life-like characters. Furthermore, as several researchers have argued, a character's visual appearance has a strong influence on the expected character behavior as well as the anticipated expertise of the character in the particular scenario it is put to use.

- **Identification of a character's role as a possible limitation on its believability**

  Based on the aforementioned analysis, we argued that the role assigned to a life-like character is another limiting factor with respect to the character's overall believability. Most of today's life-like character implementations realize some kind of virtual expert which offers advice in a very limited domain. Users interacting with these characters immediately understand this inherent limitation in the characters knowledge and corresponding behavior. Furthermore, users of such life-like character systems don't expect these characters to be anything more than an expert or servant. Freeing a character from such a limited role would hence open up new ways to increase the character's believability by using the same character not for a single, limited purpose but instead applying it in many different scenarios and for varying purposes.

- **A new approach towards multi-purpose life-like characters**

  The Migrating Character technology addresses the problem of the limited roles of life-like characters and the resulting decreased believability of the characters by realizing a multi-purpose character concept. Instead of using a special life-like character in each different application scenario, a single Migrating Character may serve as an assistant to the user in many different application scenarios in a coherent way. To guarantee for a maximal flexibility of the character implementations, device-independent technologies are used. For example, XML encoded text is used to specify character actions and for the character animations Macromedia Flash MX is used which allows for playback on most of today's operating systems and computer platforms. Furthermore, different layouts for the same Migrating Character allow to use it on different devices with varying available screen space. While keeping its general personality and appearance, a Migrating Character is capable of adapting to different scenarios by updating its knowledge base according to the application scenario and also by visibly reflecting the actual scenario in its own appearance (i.e. by wearing special clothes or by using a particularly fitting artifact). Hence, the character does not necessarily posses a vast general knowledge, but instead it may offer relevant information according to the actual scenario. Such a multi-purpose character also bears the benefit of being capable to adapt to the user's needs and preferences over a long period of time. In this way, the quality and reliability of the derived user model is improved as compared to a short term user model used for single-purpose characters.

- **Life-Like character technology entering the realm of social settings**

  Life-Like character technology basically tries to mimic human behavior. One aspect of particular interest to the corresponding research area is that of imitating social interaction among humans. However, the setup of most character implementations does not allow to integrate the character into a social setting, since they are restricted to a particular, usually stationary and hence fixed device. The social protocols which apply to these special setups for communication between characters and humans are rather simple. In a real life setting however, a life-like character may reveal its full social capabilities. It may react to varying situations, like for example when interacting with a user in a public setting as opposed to a private one. The Migrating Character technology is a first step in this direction. By allowing developers to implement life-like characters which are potentially available everywhere and in every situation, the Migrating Characters allow to explore and realize new social capabilities of life-like characters in real life, social settings. Migrating Character may eventually become electronic companions for human users.

In addition to these contributions to the research area of life-like characters, the Migrating Character concept also addresses problems and limitations typically found in mobile computing applications. In combining mobile and stationary devices into a single scenario and by introducing life-like characters in such setups, the following results were achieved:

- **Overcoming the limitation of small screen space on mobile devices**

  The inherent limitation of each mobile device is its the small screen space available. One way of dealing with this problem is to develop displays of better quality due to higher resolutions. A natural limit in this direction is however constituted by the human visual perception system. We presented a different approach by integrating stationary presentation systems into the mobile application scenario. Whenever necessary, high quality multimedia presentations may be relayed to these stationary systems instead of presenting them on the user's mobile device. The decision whether a certain multimedia object should be shown on a stationary device depends both on the properties of the multimedia data as well as on the availability of a suitable presentation device in the vicinity of the user. As part of the Migrating Character technology, we presented methods to detect suitable devices integrated in the environment. These methods also take into account personal preferences of users, when deciding for example, whether to show a specific multimedia object in high quality on a device located in another room (which demands a user to cover a certain distance) or in lower quality on the mobile device of the user. In case the user's preferences indicate that a relocation towards a stationary presentation system is acceptable, the Migrating Character technology offers efficient ways to guide the user to the desired destination in front of the stationary device.

- **Overcoming the limitation of low computational power in mobile devices**

  The Migrating Character concept, due to the complexity of the realized features, demands the use of additional stationary devices within the mobile computing scenario. By means of wireless communication, computing intensive processes may be relayed to stationary, high power computers running corresponding services. This process relay mechanism allows for services realized on a mobile platform which otherwise would be impossible to run on today's mobile devices. However, due to the often unreliable communication via wireless communication channels, a backup mechanism for situations in which the communication fails, needs to be provided. One example discussed in this work is that of natural language synthesis for the Migrating Characters. Using the discussed synthesis approach allows for a constant speech quality on both mobile and stationary devices. In case the speech synthesis service fails, the backup mechanism provided is that of using written text for the Migrating Character's utterances instead.

Instrumented environments, combining stationary and mobile devices into a single application scenario, offer not only new technology but also demand new interaction metaphors. Since user interactions with instrumented environments typically involve the use of several different services as well as devices, it is important to support users in these complex interaction setups by means of a general, coherent user interface. Furthermore, since users are interacting in physical space, a mechanism needs to be found in order to produce necessary spatial cues to support the user in finding relevant objects and devices. Throughout this work, we discussed the following solutions to the above stated problems:

- **Migrating Characters represent a coherent interface for mobile and stationary devices**

  A Migrating Character, regardless of the device it is used on or the context of its application, always appears in a coherent way. This is both true for its visual appearance as well as for the characters voice and especially for the characters behavior. Since the Migrating Character technology allows for the use of these characters on basically any presentation device integrated in an instrumented environment, they hence serve as a coherent interface metaphor for all the different devices they are used on and also for the many different services which may be accessed in the instrumented environment. Furthermore, the Migrating Characters have proven to effectively guide a user's focus between different types of devices, when using those devices in a parallel fashion.

- **Migrating Characters may be used on the majority of today's common hardware platforms and operating systems**

  Data formats as well as the software development tools used during the realization of the exemplary Migrating Character implementations have been carefully chosen in order to provide for maximum portability. Using XML formatted text files as the general underlying data structure does not limit the use to a particular operating system or specific programming languages. The XML standard is well defined and widely spread and thus provides a solid basis for the necessary data to be stored as well as for the information to be exchanged between the different components of the exemplary Migrating Character implementations. All server components have been written in Java and may hence be used on a large variety of different devices with varying operating systems. Macromedia Flash MX was chosen in all of the Migrating Character examples to realize the visual appearance of the character but also to provide the graphical user interface of the underlying application. The main benefit of Flash, apart from its advanced multimedia capabilities, is the availability of the Flash player component for the majority of today's popular hardware platforms including mobile devices as well as stationary ones. It is hence possible to develop a single Migrating Character and then use it on a variety of different devices.

- **Migrating Characters may perform unique references to physical and virtual objects**

  The Migrating Characters, similar to life-like characters inhabiting virtual worlds, are capable of performing references to real physical objects. In order to do so, the Migrating Character concepts demands the implementation of a world knowledge base for the character as well as the use of sensors which allow to detect positions of users and possibly also positions of objects. In this work, we introduced several different methods for performing references to both physical and virtual objects with Migrating Characters. Furthermore, we introduced mechanisms which allow a Migrating Character to automatically decide which referencing method is to be used, in order to disambiguate references in an arbitrary situation.

8.2 Opportunities for Further Research

The main focus of the work at hand was on the development of a concept to allow for the realization of life-like characters which should work on both mobile and stationary systems in a consistent way. The resulting Migrating Character concept explores possibilities and defines necessities for the development of life-like characters featuring the aforementioned capability.

While the conceptual part of the work tried to cover the Migrating Character technology in all its facets, each of the implemented Migrating Character prototypes could only realize subparts of the concept. Due to the complexity and extensibility of the Migrating Character concept, even all of these prototypes together do not implement every aspect of the concept. It is hence desirable to further explore the feasibility of those parts of the concept which were not yet implemented. While the discussed prototypes mainly focused on the realization of the character engine from a technical point of view, in a next step one should put a stronger accent on the communication between users and the Migrating Characters. While the PEACH prototype (see Section 6.1) realizes communicative skills of the Migrating Character by means of synthesized speech in conjunction with according lip movements, facial expressions and gestures of the character, it does not support an equally natural input method for the user.

To support coherent multimodal interaction with a Migrating Character in a mixed device scenario including both mobile and stationary devices, using a remotely run service on a stationary device, similar to the speech synthesis approach realized in PEACH (see Subsection 6.1.2.4), seems to be most promising. Realizing the speech recognition solely on the mobile platform would yield lower recognition quality due to the limited computational power of today's mobile hardware. Huang et al., 2001 discuss a technique which allows to record an audio signal via the integrated microphone on a modern PDA. Instead of directly analyzing this data, the audio signal is send as a stream via a wireless connection to a stationary server. The server may then run a far more sophisticated speech recognition algorithm on the received audio stream. As a result, the server sends back the recognized utterance to the PDA.

Speech recognition is however only a first step towards a natural communication between Migrating Characters and the users. Supporting gesture recognition in addition will demand further integration of technical devices, such as cameras for example. Whether these cameras should be attached to the mobile device or better be integrated in the environment is an open question which needs to be further investigated. A very interesting approach towards gesture recognition is discussed in (Wasinger & Wahlster, 2006). The authors define different classes of gestures. While the so-called intra-gestures refer to pointing gestures on a touch screen, extra-gestures involve the physical manipulation of objects by picking them up or putting them down. The described technology also allows for modality combinations, for example speech input and extra-gesture. Integrating this approach in the Migrating Character technology would yield new interaction possibilities.

The user study presented in Chapter 7 focused on a very specific problem addressed by the Migrating Character technology, namely the support for heterogeneous user groups using a single stationary presentation system and the attentional focus guidance during presentations spanning over mobile and stationary devices. Further evaluations need to investigate questions regarding for example the effectiveness of the physical object references performed by Migrating Characters or the general acceptance of the life-like character technology on mobile systems.

The technologies presented in this work which allow for the movement of the Migrating Characters in physical space do not give these characters the same degrees of freedom which apply to humans. While carried on a PDA, a Migrating Character is moving through three-dimensional space but this movement is only passive. All techniques for active Migrating Character locomotion developed so far do limit the characters movements to specific locations or areas. Furthermore, to allow a Migrating Character to move through physical space using the introduced technologies demands a technologically well equipped environment and specific services. So the vision of a Migrating Character being able to move everywhere and to appear at any location (like the people in Startrek when they "beam" from one area to another) does still belong to the realm of science fiction. Other ideas like, for example, a Migrating Character located on a stationary display in a shopping mall, advertising a specific product and being capable of jumping on a visitors mobile phone to guide him to the product, could possibly be realized in the near future.

A Questionnaire on Parallel Presentations on Public and Private Displays

This appendix summarizes the complete questionnaire of the user study described in Section 7.

The following explanations relate to the questionnaire for PDA1. Question number 0 asks subjects to state briefly what might distract them while trying to follow presentations. The subsequent questions on the first page translate as follows (questions one to eight have a range from 1 [I totally Agree] to 10 [I totally Disagree]):

1. I have problems remembering numbers
2. I have problems remembering data related to images
3. I can remember everything, even if I see it only once
4. I cannot concentrate on a particular task for a long period of time
5. I cannot concentrate if there are many people in the same room with me
6. I need some time to concentrate on something new
7. Background noise negatively influences my concentration
8. I know how to use a PDA

These initial questions are followed by an introduction to the first experiment phase, telling the subject how the experiment will proceed and to concentrate on the Comic character which will guide the subjects focus. In addition, subjects are told not to turn around the page, because on the next page starts already the questionnaire regarding the first presentation phase. That questionnaire is again concluded with an instruction for the subjects, telling them that there will be no visual focus during the next presentation phase. Again, the last sentence instructs the subject not to turn the page. This is just the same after the second questionnaire (in this case, the user focus will be guided by an animated symbol). The last questionnaire is followed by an additional page asking questions regarding the performance of the different signaling methods. The last two questions ask, which signaling method was preferred (and why) and whether anything during the experiment appeared particularly negative.

The questionnaires for PDAs 2 and 3 looked just the same, except for the fact that instructions regarding presentations were different. This is due to the fact that each PDA used a different signaling method (during each presentation phase) to warn subjects prior to a device switch within the presentation.

Datum:________Uhrzeit:_______Name/Vorname:_____________________________PDA1
Geschl.: m[ ] w[ ]   Alter:_____

Fragebogen

0. Bitte geben Sie in Stichworten an, wovon Sie sich waehrend einer Praesentation ablenken lassen bzw. gestoert fuehlen.

1. Ich habe Probleme mir Zahlen zu merken.
Ich stimme nicht zu 1 2 3 4 5 6 7 8 9 10 Ich stimme zu

2. Ich habe Probleme mir Daten zu Bildern zu merken.
Ich stimme nicht zu 1 2 3 4 5 6 7 8 9 10 Ich stimme zu

3. Wenn ich etwas einmal sehe kann ich mich wieder daran erinnern.
Ich stimme nicht zu 1 2 3 4 5 6 7 8 9 10 Ich stimme zu

4. Ich kann mich nicht lange auf etwas konzentrieren.
Ich stimme nicht zu 1 2 3 4 5 6 7 8 9 10 Ich stimme zu

5. Ich kann mich nicht konzentrieren wenn ich mit mehreren Personen in einem Raum bin.
Ich stimme nicht zu 1 2 3 4 5 6 7 8 9 10 Ich stimme zu

6. Ich brauche etwas Zeit um mich auf etwas Neues zu konzentrieren!
Ich stimme nicht zu 1 2 3 4 5 6 7 8 9 10 Ich stimme zu

7. Hintergrundgeraeusche stoeren meine Aufmerksamkeit.
Ich stimme nicht zu 1 2 3 4 5 6 7 8 9 10 Ich stimme zu

8. Ich bin vertraut im Umgang mit Personal Digital Assistants (PDAs).
Ich stimme nicht zu 1 2 3 4 5 6 7 8 9 10 Ich stimme zu

Einleitung:

Sie werden im folgenden Versuch Ausschnitte aus einem Dokumentarfilm der National Geographic Society (NGS) zu sehen bekommen. Die Ausschnitte befassen sich jeweils mit einem oder mehreren Forschern, Entdeckern oder Abenteurern, die von der NGS gefoerdert wurden. Die Filme laufen auf dem grossen Flachbildschirm ab. Sie erhalten zusaetzliche Informationen auf dem PDA. Die Praesentationen auf dem PDA werden automatisch gestartet, waehrend die Filme auf dem Flachbildschirm laufen. Die praesentierten Informationen (Inhalte) auf dem Flachbildschirm und dem PDA koennen sich ueberschneiden. Sobald eine Praesentation auf dem PDA startet, richten Sie bitte Ihre **volle (!!!)** Aufmerksamkeit auf die Praesentation auf dem PDA. Endet die Praesentation auf dem PDA, richten Sie Ihre Aufmerksamkeit wieder auf den Flachbildschirm. Um Ihnen den Wechsel vom einen zum anderen Geraet zu erleichtern, wird Ihnen jeweils ein Signal gegeben, kurz bevor Sie Ihre Aufmerksamkeit vom einen auf das Andere Geraet lenken muessen. Nach jeder Praesentation (also Kombination aus PDA- und Flachbildschirm-Praesentation, Dauer ca. 4 Minuten) bekommen Sie einen Fragebogen. Ihre Aufgabe besteht darin, sich so viele Informationen wie moeglich zu merken.

In der nun folgenden Praesentation achten Sie bitte (zusaetzlich zum Inhalt) auf die Comic-Figur, die Ihnen, indem Sie vom Flachbildschirm verschwindet und auf dem PDA auftaucht (und umgekehrt), Signalisiert, wohin Sie Ihre Aufmerksamkeit lenken sollen (Erlaeuterung durch Versuchsleiter).

Stop! (bitte nicht umblaettern)

Wer ist das und was macht er auf dem Bild und was sagt er?

Wer sind die beiden Personen auf dem Bild, was wird getan und was wird gesagt?

Wer/was ist das auf dem Bild zu sehen und was ist seine/ihre Aufgabe?

Wen transportiert dieses Schiff, woher kommt es und wohin segelt es?

Was wurde von Howard Carter 1922 gefunden und wo wurde es gefunden?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wie weit zurueck in die Vergangenheit wurde dem Publikum eine Reise geboten?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Mit wie viel Dollar wurde Richard Byrd’s versuch von der National Geograhpic Society gesponsert?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was soll in der Antarktis eingerichtet werden und was ist Byrd’s plan?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Zu welcher Jahreszeit will Byrd den Flug wagen?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Die Expedition vergroessert sich. Wie viele Hunde gehoeren vorher und nachher zur Mannschaft?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Aus wie vielen Maennern besteht der Vortrupp und wie lange werden sie unterwegs sein?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was moechte Byrd am Suedpol zurueck lassen?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wieso kann man die Abenteuer von Byrd miterleben?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was markiert fuer immer das aeusserste Ende der Welt?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was hat 1926 Byrd mit seinem Freund Floyd Bennet schon getan?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was tut Byrd 1929 mit 42 Leuten ?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was sehen die Expeditionsteilnehmer vom 18 April bis 23 August nicht?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was wird abgeworfen damit genuegend Hoehe erreicht wird um die Bergketten zu ueberfliegen?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

In welchem Zustand erleben Byrd und Kameraden den triumphalen Empfang im Lager?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wie lange dauerte der Flug?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Erlaeuterung zum naechsten Abschnitt:

In der nun folgenden Praesentation bekommen Sie kein Signal, die Praesentation auf dem PDA startet und endet ohne Vorwarnung. Bitte versuchen Sie dennoch, Ihre Aufmerksamkeit so schnell wie moeglich vom Einen auf das Andere Geraet zu lenken.

Stop! (bitte nicht umblaettern)

Wer ist das und welcher Expedition gehoert er an?

Wer ist das und welcher Expedition gehoert er an?

Welches Schiff ist das und zu welcher Expedition gehoert es?

Was ist das und welchen Zweck erfuellt es?

Vor welcher Mittelmeerkueste wird ein antikes Schiff erkundet und vor wie vielen Jahren ist es gesunken?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Welche Art von Erkenntnissen ueber das Antike Schiff werden von den Forschern gewonnen?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

In welcher Tiefe und wie weit entfernt von der Kueste liegt das Schiffswrack?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wie lang war das Schiff (in Meter)?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Aus mindestens wie viel verschiedenen Kulturkreisen stammten die Waren an Bord?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Gehoert Elfenbein z u den Waren?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was war ein weiterer Beweis dafuer, dass das Schiff aus der Bronzezeit stammt?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Stimmt es das die Phoenizischen Schiffe aelter sind als dieses Schiff?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wo spuerte Ballard die Wracks der phoenizischen Schiffe auf!

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wann und wodurch wurde das britische Passagierschiff Lusitania versenkt?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wann(in welchem Jahr) sank die Titanic?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wann (in welchem Jahr) findet Ballard das Wrack der Titanic?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wie viele Jahre spaeter findet eine 2. Expedition zur Titanic statt?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was sagt man, habe Robert Ballard unter Wasser?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was entdeckt Robert Ballard 1977 am Grund des Ostpazifik?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Welche Daten wurden ueber das Schiff Bismarck praesentiert?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

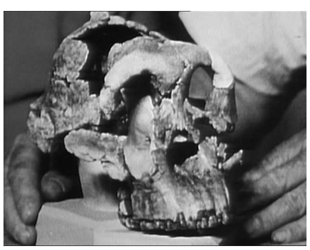
Erlaeuterung zum naechsten Abschnitt:

In der nun folgenden Praesentation achten Sie bitte (zusaetzlich zum Inhalt) auf das Symbol der National Geographic Society (oben links auf dem Flachbildschirm bzw. in der Mitte des PDA Bildschirmes), das Ihnen, in Form einer Animation Signalisiert, auf welches Geraet Sie Ihre Aufmerksamkeit lenken sollen. Die Animation startet jeweils auf dem Geraet, auf das Sie sich als naechstes konzentrieren sollen (z.B. auf dem PDA, wenn Sie gerade eine Praesentation auf dem Flachbildschirm verfolgen). Sobald die Animation beendet ist, startet die Praesentation auf demselben Geraet. Achtung: Die Animation wird nicht wiederholt!

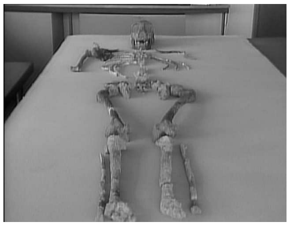
Stop! (bitte nicht umblaettern)

Was ist das?

Wer ist das?



Wessen Schaedel ist das, wo und von wem wurde er gefunden?



Wessen Knochen sind das, wo und von wem wurden sie gefunden?

Woraus besteht der Boden in der Olduway Schlucht?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wie alt sind die Ablagerungen der Erschichten in der Olduway Schlucht?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was wollten Louis und Mary Leakey herausfinden?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wie lange ernteten die Leakeys Skepsis?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was waren die wesentlichen Annahmen der Leakey’s ueber den Menschen?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was begann Jane Goodall auf draengen von Louis Leakey?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was fuer Gemeinsamkeiten werden von Jane Goodall beobachtet?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Eine andere Schuelerin der Leakeys widmet sich des Studium der Berggorilla in Ruanda.
Was studiert eine der Schuelerinnen Leakey’s in Ruanda?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Womit gelingt es Diane Fossey vertrauen zu erwerben?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wie heisst eines der Lieblingstiere von Diane Fossey?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Welches Schicksal erleidet eines der Lieblingstiere von Diane Fossey und warum?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Welches Schicksal erleidet Diane Fossey?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wo suchen Louis Leakey und seine Frau nach den Spuren des Vormenschen?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Welchen Teil des Schaedels finden Louis Leakey und seine Frau 1959 als erstes in der Olduway Schlucht?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Welche Lehrmeinung wird durch diese Entdeckung widerlegt?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Was findet Mary Leakey in der Naehe der Olduway Schlucht?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wo findet Richard Leakey Knochenreste?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Wie alt ist Richard Leakey zu dieser Zeit?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

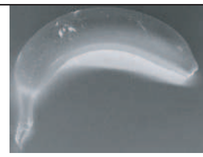
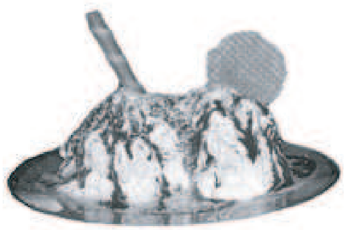
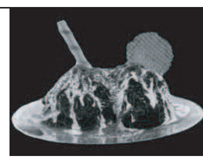
Stimmt es das die Knochen die von Richard Leakey gefunden wurden zur selben Gattung gehoeren wie der Schaedel, der von seinen Eltern gefunden wurde?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Stimmt es das mit diesem Fund klar wurde, aus welcher Gattung sich der Homosapien entwickelt hat?

Loesung:	Aus der Praesentation erfahren!	Vorwissen	Geraten

Zusatzfragen:

Haben Sie dieses Bild gesehen?	Haben Sie auch diese Bilder gesehen? (pro Zeile nur maximal ein Kreuz)
 ja nein	  
 ja nein	  
 ja nein	  

Welche Signalisierungsmethode hat Ihnen am besten gefallen und wieso?

Ist Ihnen irgendetwas waehrend der Praesentationen besonders negativ aufgefallen?

B Exemplary Presentation Script for the PEACH Museum Guide

The XML-based script on the next page is an example for a presentation script used in the PEACH museum guide project. Each such script in PEACH is related to a single station or exhibit within the museum. The example given is related to a fresco related to the month of February in the Torre Aquila of the Castello del Buon Consciglio in Trento, Italy. The scripts are designed as follows:

Each script starts with the “PRESENTATION” tag which resembles a header part for the script. In this part, the station’s title and an initial image to be shown are defined. Whenever an image is referred to within the script, further options are available. While the “zoom” option allows the content designer to define a region of the given image to be shown, instead of fitting the whole image on the screen of either mobile or stationary device, the detail option allows to highlight a specific area of the image. In order to do so, most of the image display region is overlaid by a dark, partly transparent layer. Only the specified region (e.g. upper left corner, lower middle region etc.) is left untouched and hence becomes very striking.

Each script is subdivided into several regions, in the given example these are “INTRO”, “TECHNICAL” and “HELP”. These regions correspond to the options given to the user while interacting with the device.

Each region is composed of several parts which are played back as a sequence when the corresponding information is requested by the user. Each part consists of the text to be spoken, the name of the character who is going to speak, the according gesture, and an image definition similar to the one in the header part. In addition, for each image, the “transition option” defines, how the active image should be replaced with the new one (e.g. blend, cut etc.).

```
<PRESENTATION title="February" image="february.jpg"
              size="402x585" zoom="ALL" detail="AP">
  <INTRO>
    <PART
       text="At the top there are about twenty young women (whose
             faces have been visibly repainted) watching a
             tournament going on below the curtain walls. Four
             knights are jousting against four others. Pages and
             servants are helping the knights to dress or are
             picking up the pieces of broken weapons that are
             on the ground."
       characterName="diva"
       gesture="showUpLeft_talk"
       image="february.jpg"
       size="402x585"
       zoom="0x0-402x350"
       transition="CUT"
       detail="AP">
    </PART>
    <PART
       text="At the bottom on the right is a blacksmith’s
             workshop, a plebeian antithesis to the tournament
             going on in the upper part of the painting which
             is chiefly an aristocratic activity."
       characterName="diva"
       gesture="talkFrontal"
       image="february.jpg"
       size="402x585"
       zoom="250x100-402x400"
       transition="CUT"
       detail="LRP">
    </PART>
    <PART
       text="The choice of a tournament for the month of February
             is related to the jousts and revelries that took
             place in carnival time. The picture of young women
             behind a parapet is often illustrated in the French
             tapestries of the 1300’s and is similar to a
             decoration in the Runkelstein Castle near Bolzano."
       characterName="diva"
       gesture="showUpLeft_talk"
       image="february.jpg"
       size="402x585"
       zoom="ALL"
       transition="BLEND"
       detail="AP">
    </PART>
  </INTRO>
```

```
<TECHNICAL>
  <PART
     text="The february fresco measures 1 point seventy meters
           on 2 point 22 meters and is subdivided into six days."
     characterName="diva"
     gesture="attentionCalling_talk"
     image="february.jpg"
     size="402x585"
     zoom="ALL"
     transition="BLEND"
     detail="AP">
  </PART>
</TECHNICAL>
<HELP>
  <PART
     text="I will give you a brief introduction on the
           information system. Spread over the entire site,
           there are several information booth installed."
     characterName="diva"
     gesture="lookLeft_talk"
     image=""
     size="500x713"
     zoom="ALL"
     transition="CUT"
     detail="AP">
  </PART>
  <PART
     text="Whenever you reach one of these booth, the name
           of the booth is displayed in the text field
           located at the bottom of the screen."
     characterName="diva"
     gesture="showLeft_talk"
     image=""
     size="500x713"
     zoom="ALL"
     transition="CUT"
     detail="AP">
  </PART>
  <PART
     text="Furthermore a short introduction to the booth
           is automatically started. By pressing one of the
           Buttons: Intro or Technical, you can request
           detailed information on each subject."
     characterName="diva"
     gesture="showUpLeft_talk"
     image=""
     size="500x713"
     zoom="ALL"
     transition="CUT"
```

```
       detail="AP">
    </PART>
    <PART
       text="You may stop a presentation or skip forward by
             pressing the buttons next to the progress bar."
       characterName="diva"
       gesture="lookLeft_talk"
       image=""
       size="500x713"
       zoom="ALL"
       transition="CUT"
       detail="AP">
    </PART>
  </HELP>
</PRESENTATION>
```

Prior to sending the script to the desired presentation device, the server will parse the script itself. In particular, it will read the texts defined in the script and transform them via speech synthesis into MP3 encoded audio files (as described in section 6.1.2.4). Once the speech synthesis is done, the server will substitute the texts with the corresponding audio filenames. The following is the same presentation script as above, however it reflects the changes done by the server.

```
<PRESENTATION title="February" image="february.jpg"
              size="402x585" zoom="ALL" detail="AP">
  <INTRO>
    <PART
       text="bf8314831d62600ea5a6.mp3"
       characterName="diva"
       gesture="showUpLeft_talk"
       image="february.jpg"
       size="402x585"
       zoom="0x0-402x350"
       transition="CUT"
       detail="AP">
    </PART>
    <PART
       text="92087d369744e539df40.mp3"
       characterName="diva"
       gesture="talkFrontal"
       image="february.jpg"
       size="402x585"
       zoom="250x100-402x400"
       transition="CUT"
       detail="LRP">
    </PART>
    <PART
       text="df59053cc4f5cf3101e4.mp3"
```

```
      characterName="diva"
      gesture="showUpLeft_talk"
      image="february.jpg"
      size="402x585"
      zoom="ALL"
      transition="BLEND"
      detail="AP">
   </PART>
 </INTRO>
 <TECHNICAL>
   <PART
      text="ceeb6fbf68f0dcc8ee4b.mp3"
      characterName="diva"
      gesture="attentionCalling_talk"
      image="february.jpg"
      size="402x585"
      zoom="ALL"
      transition="BLEND"
      detail="AP">
   </PART>
 </TECHNICAL>
 <HELP>
   <PART
      text="af30f8ce5d1cdb05fd43.mp3"
      characterName="diva"
      gesture="lookLeft_talk"
      image=""
      size="500x713"
      zoom="ALL"
      transition="CUT"
      detail="AP">
   </PART>
   <PART
      text="6b050e70f3bdb1cca469.mp3"
      characterName="diva"
      gesture="showLeft_talk"
      image=""
      size="500x713"
      zoom="ALL"
      transition="CUT"
      detail="AP">
   </PART>
   <PART
      text="46b948e230ba16cce208.mp3"
      characterName="diva"
      gesture="showUpLeft_talk"
      image=""
      size="500x713"
      zoom="ALL"
```

```
       transition="CUT"
       detail="AP">
    </PART>
    <PART
       text="f4b98cfc39c8c1cad8af.mp3"
       characterName="diva"
       gesture="lookLeft_talk"
       image=""
       size="500x713"
       zoom="ALL"
       transition="CUT"
       detail="AP">
    </PART>
  </HELP>
</PRESENTATION>
```

C Exemplary Video Script for the PEACH Museum Guide

The following is an exemplary XML-based video presentation script which was automatically generated by the video composer component of the PEACH project. The script is subdivided into two main sections, namely the shots definition and the editing part. While the shots definition part defines each shot which is to be used while rendering the video, the editing part is used to define the sequence of the previously defined shots. The editing part controls also the way in which shots are concatenated. In the given example, the first shot uses the condition “display” which means that there will be no transition effect. The second shot uses the crossfade transition, meaning that shot01 will be faded out and shot02 will be faded in simultaneously.

Each shot definition in the shots section consists of two parts, namely the video-track definition and the corresponding audio-track definition. The video-track part holds instructions which control camera movements over an image which is defined in the header of the shot. The audio-track part defines a sequence of audio files and pauses synchronized with the video-track definitions. As opposed to the solution for the Migrating Character in PEACH, the video clip generation uses pre-synthesized or recorded speech in audio files, since these had to be annotated with time stamps. Apart from the image to be used, the header of each shot also holds information about the general topics addressed by this shot as well as a perspective tag which is used by the user model in PEACH.

Finally, each movie has also a header which defines the perspective and topic of the whole video-clip.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
  <movie perspective="introductory" topic="january">
    <shots>
        <shot id="shot01" image="jan-scaled.jpg"
              topic="january,snowball-fight,castle"
              perspective="introductory">
                <video-track>
                        <pause duration="8"/>
                        <zoom scale="1" duration="3"/>
                        <pause duration="3"/>
                        <move x="100" y="120" duration="3"/>
                        <pause duration="12"/>
                </video-track>
                <audio-track>
                        <play audio="january.mp3"/>
                        <audio-pause duration="0.5"/>
                        <play audio="snowball-fight.mp3"/>
                        <audio-pause duration="1.5"/>
                        <play audio="castle.mp3"/>
                        <audio-pause duration="0.5"/>
                        <play audio="castle2.mp3"/>
                        <audio-pause duration="0.5"/>
                </audio-track>
        </shot>
        <shot id="shot02" image="windows1.jpg" topic="windows">
                <video-track>
                        <pause duration="5"/>
                </video-track>
                <audio-track>
                        <play audio="win1.mp3"/>
                </audio-track>
        </shot>
    </shots>
    <editing>
        <display shot="shot01"/>
        <crossfade shot="shot02"/>
    </editing>
</movie>
```

D Exemplary Presentation Script for the VRI

The XML-based script on the next page is an example of a controlling script for the Virtual Room Inhabitant (VRI). The script consists of several subscripts which may be played either in sequence or in any other given order. Each of these subscripts may either describe actions to be performed by the character or character transitions. In the first case, a sequence of parts, each consisting of a gesture definition for the character and the name of a corresponding mp3 file holding the aural information which is to be performed by the character in conjunction with the defined gesture, is described. As opposed to the PEACH project, the texts to be spoken by the VRI character were pre-synthesized. The speech synthesis mechanism used in PEACH could be easily incorporated into the VRI setup, however due to limited computational resources, the demo setup did not include the ”real time speech synthesis”.

In case of a character transition, two different methods may come to use. The first one is a smooth character movement from its current location to a target destination. The movement is defined in the “move” tag and uses a point in three dimensional space and a space of time, defined in milliseconds, as its parameters. These parameters are interpreted by the VRI server which sends corresponding commands to the spatial audio device, to the steerable projector and to the VRI character animation (as described in detail in Sections 6.2.3, 6.2.4, 6.2.5 and 6.2.6). The second method for a character transition is a jump. A jump does not relocate the VRI by moving it smoothly to its new location, but instead the character disappears at its current location (in case the character is visible at all in that moment) and it reappears at another location in the room. The parameters of the “jump” tag vary from those of the “move” tag. Instead of a single point in three dimensional space, the “jump” tag requires the definition of four such points. Each point defines one corner of a so-called virtual display (see section 6.2.3) around the location where the character should appear.

```
<VRI-script>
    <subscript>
        <part gesture="Hips" sound="hello.mp3"/>
        <part gesture="LookFrontal" sound="shoppingcart.mp3"/>
        <part gesture="swirl" sound="SoundLoop.mp3"/>
    </subscript>
    <subscript>
        <move x="0.0" y="2.3" z="-4.35" time="8000"/>
    </subscript>
    <subscript>
        <jump llx="-0.2" lly="2.0" llz="-4.47" lrx="0.4"
              lry="2.0" lrz="-4.47" urx="0.4" ury="2.6"
              urz="-4.47" ulx="-0.2" uly="2.6" ulz="-4.47"
              time="0"/>
    </subscript>
    <subscript>
        <move x="3.65" y="2.3" z="-4.47" time="10000"/>
    </subscript>
    <subscript>
        <jump llx="3.8" lly="2.0" llz="-4.75" lrx="3.8"
              lry="2.0" lrz="-4.15" urx="3.8" ury="2.6"
              urz="-4.15" ulx="3.8" uly="2.6" ulz="-4.75"
              time="0"/>
    </subscript>
    <subscript>
        <move x="3.8" y="2.3" z="-3.0" time="5000"/>
    </subscript>
    <subscript>
        <move x="3.8" y="1.4" z="-3.0" time="5000"/>
    </subscript>
    <subscript>
        <part gesture="LookBehind" sound="panel.mp3"/>
        <part gesture="LookFrontal" sound="tryout.mp3"/>
    </subscript>
</VRI-script>
```

List of Figures

List of Tables

Bibliography

Alibali, M. W., & Goldin-Meadow, S. (1993). Modeling Learning Using Evidence From Speech and Gesture. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (p. 203-208). Hillsdale, USA: Lawrence Erlbaum Associates.

Allbeck, J. M., & Badler, N. I. (2001). Towards Behavioral Consistency in Animated Agents. In *DEFORM 2000/AVATARS 2000: Proceedings of the IFIP TC5/WG5.10 DEFORM'2000 Workshop and AVATARS'2000 Workshop on Deformable Avatars* (p. 191-205). Deventer, The Netherlands: Kluwer, B.V.

André, E., Dorfmüller-Ulhaas, K., & Rehm, M. (2005). Engaging in a Conversation with Synthetic Characters Along the Virtuality Continuum. In *Proceedings of the Fifth International Symposium on Smart Graphics* (Vol. 3638, p. 1-12). Frauenwörth Cloister near Munich, Germany: Springer LNCS.

André, E., Dorfmüller-Ulhaas, K., & Rist, T. (2004). Embodied Conversational Characters: Wandering Between the Digital and the Physical World. In W. Wahlster (Ed.), *Special Journal Issue "Conversational User Interfaces", IT - Information Technology* (Vol. 46, p. 332-340). München, Germany: Oldenbourg.

André, E., Rist, T., & Müller, J. (1997). The PPP Persona: Towards a Highly Personalized User Interface. In C. Stephanidis (Ed.), *UI4All 1997: Proceeding of the Third ERCIM Workshop on User Interface for All* (p. 9-17). Obernai, France: INRIA Lorraine.

André, E., Rist, T., & Müller, J. (1998a). Integrating Reactive and Scripted Behaviors in a Life-Like Presentation Agent. In K. P. Sycara & M. Wooldridge (Eds.), *Proceedings of the Second International Conference on Autonomous Agents (Agents 1998)* (p. 261-268). Minneapolis/St.Paul, USA: ACM Press.

André, E., Rist, T., & Müller, J. (1998b). WebPersona: A Life-Like Presentation Agent for the World-Wide Web. *Knowledge-Based Systems*, *11*(1), 25-36.

Andrè, E., Rist, T., Mulken, S. van, Klesen, M., & Baldes, S. (2000). The Automated Design of Believable Dialogues for Animated Presentation Teams. In *Embodied Conversational Agents* (p. 220-255). Cambridge, USA: MIT Press.

Aoki, P. M., Grinter, R. E., Hurst, A., Szymanski, M. H., Thornton, J. D., & Woodruff, A. (2002). Sotto Voce: Exploring the Interplay of Conversation and Mobile Audio Spaces. In *CHI 2002: Proceedings of the SIGCHI Conference on Human Factors In Computing Systems* (p. 431-438). New York, USA: ACM Press.

Arons, B. (1992). A Review of The Cocktail Party Effect. *Journal of the American Voice I/O Society*, *12*, 35-50.

Baldes, S., Gebhard, P., Kipp, M., Klesen, M., Rist, P., Rist, T., & Schmitt, M. (2002). The Interactive CrossTalk Installation: Meta-Theater with Animated Presentation Agents. In H. Prendinger (Ed.), *Proceedings of the International Workshop on Lifelike Animated Agents - Tools, Affective Functions, and Applications, held in conjunction with the Seventh Pacific Rim International Conference on Artificial Intelligence (PRICAI 2002)* (p. 9-15). Tokyo, Japan: Japanese Society for Artificial Intelligence (JSAI).

Bates, J. (1992). Virtual Reality, Art, and Entertainment. *Presence: Teleoperators and Virtual Environments*, *1*(1), 133-138.

Bates, J. (1994). The Role of Emotion in Believable Agents. *Communications of the ACM*, *37*(7), 122-125.

Baumann, J., Hohl, F., Rothermel, K., Strasser, M., & Theilmann, W. (2002). MOLE: A Mobile Agent System. *Software: Practice and Experience*, *32*(6), 575-603.

Baus, J. (2003). *Ressourcenadaptierende hybride Personennavigation.* Amsterdam, The Netherlands: PhD. thesis, DISKI series Vol. 268, IOS Press.

Baus, J., Krüger, A., & Butz, A. (visited 2005). *The IRREAL Project Homepage.* (http://w5.cs.uni-sb.de/irreal/)

Behr, J., Dähne, P., & Roth, M. (2004). Utilizing X3D for Immersive Environments. In *Web3D 2004: Proceedings of the ninth international conference on 3D Web technology* (pp. 71–78). Monterey, USA: ACM.

Bühler, D., Häussler, J., Krüger, S., & Minker, W. (2002). The SmartKom Mobile Multi-Modal Dialogue System. In E. André, L. Dybkjær, P. Heisterkamp, & W. Minker (Eds.), *Extended Abstracts of the ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments* (p. 35-38). Kloster Irsee, Germany: University of Southern Denmark.

Bledsoe, W. (1986). I had a Dream: AAAI Presidential Address. *AI Magazine*, *7*(1), 57-61.

Brandherm, B., & Schwartz, T. (2005). Geo Referenced Dynamic Bayesian Networks for User Positioning on Mobile Systems. In T. Strang & C. Linnhoff-Popien (Eds.), *LoCA 2005: Proceedings of the First International Workshop on Location- and Context-Awareness held in conjunction with the PERVASIVE 2005* (Vol. 3479, p. 223-234). Heidelberg, Germany: Springer LNCS.

Broadbent, J., & Marti, P. (1997). Location Aware Mobile Interactive Guides: usability issues. In *ICHIM 1997: Proceedings of the Fourth International Conference on Hypermedia and lnteractivity in Museums* (p. 88-98). Paris, France: Archives and Museum Informatics.

Brown, P. J. (1996). The Stick-e Document: A Framework for Creating Context-Aware Applications. *Origination, Dissemination, and Design (EP-ODD)*, *8*(1), 1-13.

Bruner, S., & Tagiuri, R. (1954). The Perception of People. In G. Lindzey (Ed.), *Handbook of Social Psychology* (Vol. 2, p. 634-654). Cambridge, USA: Addison- Wesley.

Butz, A., Baus, J., & Krüger, A. (2000). Augmenting Buildings with Infrared Information. In *ISAR 2000: Proceedings of the International Symposium on Augmented Reality* (p. 93-96). Munich, Germany: IEEE Computer Society Press.

Butz, A., Schneider, M., & Spassova, M. (2004). SearchLight - A Lightweight Search Function for Pervasive Environments. In A. Ferscha & F. Mattern (Eds.), *PERVASIVE 2004: Pervasive Computing, Second International Conference* (Vol. 3001, p. 351-356). Vienna, Austria: Springer LNCS.

Carmichael, D., Kay, J., & Kummerfeld, B. (2005). Consistent Modelling of Users, Devices and Sensors in a Ubiquitous Computing Environment. *User Modeling and User-Adapted Interaction*, *15*(3-4), 197-234.

Cassell, J. (2000). Embodied Conversational Interface Agents. *Communications of the ACM*, *43*(4), 70-78.

Cassell, J., Steedman, M., Badler, N., Pelachaud, C., Stone, M., Douville, B., Prevost, S., & Achom, B. (1994). Modeling the Interaction Between Speech and Gesture. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (p. 153-158). Atlanta, USA: Lawrence Erlbaum Associates.

Chen, G., & Kotz, D. (2000). *A Survey of Context-Aware Mobile Computing Research* (Tech. Rep. No. TR2000-381). Dartmouth, USA: Department of Computer Science, Dartmouth College.

Cheverst, K., Davies, N., Mitchell, K., Friday, A., & Efstratiou, C. (2000). Developing a Context-Aware Electronic Tourist Guide: Some Issues and Experiences. In T. Turner, G. Szwillus,

M. Czerwinski, & P. Fabio (Eds.), *CHI 2000: Proceedings of the Conference on Human Factors in Computing Systems* (p. 17-24). The Hague, The Netherlands: ACM.

Cheverst, K., Davies, N., Mitchell, K., & Smith, P. (2000). Providing Tailored (Context-Aware) Information to City Visitors. In P. Brusilovsky, O. Stock, & C. Strappavara (Eds.), *AH 2000: Proceedings of International Conference on Adaptive Hypermedia and Adaptive Web-based Systems* (Vol. 1892, p. 73-85). Trento, Italy: Springer LNCS.

Clement, A. (1994). Computing at Work: Empowering Action by ”Low-Level Users”. *Communications of the ACM*, *37*(1), 53-63.

Cole, R. A., Mariani, J., Uszkoreit, H., Varile, G., Zaenen, A., Zue, V., & Zampolli, A. (Eds.). (1997). *Survey of the State of the Art in Human Language Technology.* Cambridge, UK: Cambridge University Press and Giardini.

Darwin, C. (1898). *The Expression of the Emotions in Man and Animals.* Oxford, USA: Oxford University Press, 3rd edition (March, 1998).

Dautenhahn, K. (1999). Embodiment and Interaction in Socially Intelligent Life-Like Agents. *Lecture Notes in Computer Science*, *1562*, 102-142.

Davies, N., Mitchell, K., Cheverst, K., & Blair, G. (1998). Developing a Context Sensitive Tourist Guide. In C. Johnson (Ed.), *Proceedings of the First Workshop on Human Computer Interaction With Mobile Devices* (p. 64-68). Glasgow, Scotland: GIST Technical Report G98-1.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*, 193-222.

Dijk, T. A. van, & Kintsch, W. (1983). *Strategies of Discourse Comprehension.* New York: Academic Press.

Dorfmüller-Ulhaas, K., & Andre, E. (2005). The Synthetic Character Ritchie: First Steps Towards a Virtual Companion for Mixed Reality. In *ISMAR 2005: Prococeedings of the IEEE Interational Symposium on Mixed and Augmented Reality* (p. 178-179). Vienna, Austria: IEEE Computer Society Press.

Dryer, D. C. (1999). Getting Personal with Computers: How to Design Personalities for Agents. *Applied Artificial Intelligence*, *13*(3), 273-295.

Dryer, D. C., Eisbach, C., & Ark, W. S. (1999). At What Cost Pervasive? A Social Computing View of Mobile Computing Systems. *IBM Systems Journal*, *38*(4), 652-676.

Egeth, H., & Yantis, S. (1997). Visual Attention: Control, Representation, and Time Course. *Annual Review of Psychology*, *48*, 269-297.

Ekman, P. (1993). Facial Expression and Emotion. *American Psychologist*, *48*(4), 384-392.

Ekman, P., & Friesen, W. V. (1971). Constants Across Cultures in the Face and Emotion. *Journal of Personality and Social Psychology*, *17*(2), 124-129.

Elliott, C., & Brzezinski, J. (1998). Autonomous Agents as Synthetic Characters. *AI Magazine*, *19*(2), 13-30.

Eriksen, C., & Hoffman, J. (1972). Temporal and Spatial Characteristics of Selective Encoding From Visual Displays. *Perception & Psychophysics*(12), 201-204.

Eriksen, C., & Hoffman, J. (1973). The Extent of Processing of Noise Elements During Selective Coding From Visual Displays. *Perception & Psychophysics*(14), 155-160.

Fauconnier, G. (1997). *Mappings in Thought and Language.* Cambridge, USA: Cambridge University Press.

Fillmore, C. J. (1982). Towards a Descriptive Framework for Spatial Deixis. In R. J. Jarvella & W. Klein (Eds.), *Speech, Place, and Action* (p. 31-59). Chichester, UK: Wiley.

Forman, G. H., & Zahorjan, J. (1999). The Challenges of Mobile Computing. *Mobility: Processes, Computers, and Agents*, 269-285.

Franklin, D., & Flaschbart, J. (1998). All Gadget and No Representation Makes Jack a Dull Environment. In *Proceedings of the AAAI 1998 Spring Symposium on Intelligent Environments, AAAI Technical Report SS-98-02* (p. 155-160). Stanford, USA: AAAI Press.

Freksa, C., & Habel, C. (1990). Warum interessiert sich die Kognitionsforschung für die Darstellung räumlichen Wissens? In C. Freksa & C. Habel (Eds.), *Repräsentation und Verarbeitung räumlichen Wissens* (p. 1-15). Berlin, Heidelberg, New York: Springer.

Gapp, K.-P. (1997). *Objektlokalisation - Ein System zur sprachlichen Raumbeschreibung.* Wiesbaden: Deutscher Universitäts Verlag.

Göbel, S., Schneider, O., Iurgel, I., Feix, A., Knöpfle, C., & Rettig, A. (2004). VirtualHuman: Storytelling and Computer Graphics for a Virtual Human Platform. In S. Göbel, U. Spierling, A. Hoffmann, I. Iurgel, O. Schneider, J. Dechau, & A. Feix (Eds.), *TIDSE 2004: Proceedings of the Second International Conference on Technologies for Interactive Digital Storytelling and Entertainment* (Vol. 3105, p. 79-88). Darmstadt, Germany: Springer LNCS.

Golledge, R. G. (1999). Human Wayfinding and Cognitive Maps. In R. G. Golledge (Ed.), *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes* (p. 5-45). Baltimore, USA: Johns Hopkins Press.

Goodman, B. A. (1985). Repairing Reference Identification Failures by Relaxation. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics* (pp. 204–217). Morristown, NJ, USA: Association for Computational Linguistics.

Gratch, J., Rickel, J., André, E., Cassell, J., Petajan, E., & Badler, N. I. (2002). Creating Interactive Virtual Humans: Some Assembly Required. *IEEE Intelligent Systems*, *17*(4), 54-63.

Habel, C., Herweg, M., & Rehkämper, K. (1989). *Raumkonzepte in Verstehensprozessen: Interdisziplinäre Beiträge zu Sprache und Raum.* Tübingen, Germany: Niemeyer.

Haber, H., Leach, A., Schudy, S., & Sideleau, B. (1982). *Comprehensive Psychiatric Nursing (2nd edition).* New York, USA: McGraw-Hill, Inc.

Handel, S. (1989). *Listening: An Introduction to the Perception of Auditory Events.* Massachusetts, USA: MIT Press.

Hayes-Roth, B. (2004). What Makes Characters Seem Life-Like? In H. Prendinger & M. Ishizuka (Eds.), *Life-Like Characters - Tools, Affective Functions, and Applications* (p. 447-468). Berlin, Heidelberg, New York: Springer Cognitive Technologies.

Hayes-Roth, B., & Doyle, P. (1998). Animate Characters. *Autonomous Agents and Multi-Agent Systems*, *1*(2), 195-230.

Heckmann, D. (2005). *Ubiquitous User Modeling.* Amsterdam, The Netherlands: PhD. thesis, DISKI series Vol. 297, IOS Press.

Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., & Wilamowitz-Moellendorff, M. von. (2005). GUMO - the General User Model Ontology. In L. Ardissono, P. Brna, & A. Mitrovic (Eds.), *UM 2005: Proceedings of the Tenth International Conference on User Modeling* (Vol. 3538, p. 428-432). Edinburgh, UK: Springer LNCS.

Herrmann, T., & Grabowski, J. (1994). *Sprechen : Psychologie der Sprachproduktion.* Heidelberg , Berlin , Oxford: Spektrum, Akadedmischer Verlag.

Hood, M. G. (1983). Staying Away: Why People Choose not to Visit Museums. *Museum News*, *61*(4), 50-57.

Huang, X., Acero, A., Chelba, C., Deng, L., Droppo, J., Duchene, D., Goodman, J., Hon, H., Jacoby, D., Jiang, L., Loynd, R., Mahajan, M., Mau, P., Meredith, S., Mughal, S., Neto, S., Plumpe, M., Steury, K., Venolia, G., Wang, K., & Wang, Y. (2001). MiPad: A Multimodal Interaction Prototype. In *ICASS 2001: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (p. 9-12). Salt Lake City, USA: IEEE Computer Society Press.

Hull, R., Neaves, P., & Bedford-Roberts, J. (1997). Towards Situated Computing. In *ISWC 1997: Proceedings of the First International Symposium on Wearable Computers* (p. 146-153). Los Alamitos, USA: IEEE Computer Society Press.

Izard, C. (1971). *The Face of Emotion.* New York, USA: Appleton-Century-Crofts.

James, W. (1890). *The Principles of Psychology (2 vols.).* New York, USA: Henry Holt.

Johanson, B., & Fox, A. (2002). The Event Heap: A Coordination Infrastructure for Interactive Workspaces. In *WMCSA 2002: Proceedings of the Fourth IEEE Workshop on Mobile Computing Systems and Applications* (p. 83). Callicoon, USA: IEEE Computer Society Press.

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education*, *11*, 47-78.

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge, USA: Harvard University Press. Cambridge, Eng.: Cambridge University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Erlbaum.

Kawamoto, S., Shimodaira, H., Nitta, T., Nishimoto, T., Nakamura, S., Itou, K., Morishima, S., Yotsukura, T., Kai, A., Lee, A., Yamashita, Y., Kobayashi, T., Tokuda, K., Hirose, K., Minematsu, N., Yamada, A., Den, Y., Utsuro, T., & Sagayama, S. (2002). Open-Source Software for Developing Anthropomorphic Spoken Dialog Agents. In M. Ishizuka & A. Satter (Eds.), *PRICAI 2002: Proceedings of the International Workshop on Lifelike Animated Agents* (Vol. 2417, p. 64-69). Tokyo, Japan: Springer LNCS.

Kawamoto, S., Shimodaira, H., Nitta, T., Nishimoto, T., Nakamura, S., Itou, K., Morishima, S., Yotsukura, T., Kai, A., Lee, A., Yamashita, Y., Kobayashi, T., Tokuda, K., Hirose, K., Minematsu, N., Yamada, A., Den, Y., Utsuro, T., & Sagayama, S. (2004). Galatea: Open-source Software for Developing Anthropomorphic Spoken Dialog Agents. In H. Prendinger & M. Ishizuka (Eds.), *Life-Like Characters - Tools, Affective Functions, and Applications* (p. 187-211). Berlin, Heidelberg, New York: Springer Cognitive Technologies.

Kendon, A. (1994). Do Gestures Communicate? A Review. *Research on Language and Social Intraction*, *27*(3), 175-200.

Khullar, S. C., & Badler, N. I. (2001). Where to Look? Automating Attending Behaviors of Virtual Human Characters. *Autonomous Agents and Multi-Agent Systems*, *4*(1-2), 9-23.

Kiesler, S., & Sproull, L. (1997). ”Social” Human-Computer Interaction. *Human Values and the Design of Computer Technology*, 191-199.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition.* New York, USA: Cambridge University Press.

Kipp, M. (2003). *Gesture Generation by Imitation: From Human Behavior to Computer Character Animation.* Boca Raton, USA: PhD. thesis, Saarland University, Dissertation.com.

Klaus Tschira Foundation. (visited 2006). *EML - DEEPMAP website.* (URL: http://www.eml-development.de/deutsch/forschung/deepmap/index.php)

Kobsa, A., Allgayer, J., Reddig, C., Reithinger, N., Schmauks, D., Harbusch, K., & Wahlster, W. (1986). Combining Deictic Gestures and Natural Language for Referent Identification. In *Proceedings of the Eleventh Conference on Computational Linguistics* (pp. 356–361). Morristown, NJ, USA: Association for Computational Linguistics.

Koleva, B., Schnädelbach, H., Benford, S., & Greenhalgh, C. (2000). Traversable Interfaces Between Real and Virtual Worlds. In *CHI 2000: Proceedings of the SIGCHI Conference on Human factors in computing systems* (p. 233-240). The Hague, The Netherlands: ACM Press.

Kowalski, A. (visited 2006). *SmartKom website.* (URL: http://www.smartkom.org)

Kray, C., Krüger, A., & Endres, C. (2003). Some Issues on Presentations in Intelligent Environments. In E. Aarts, R. Collier, E. van Loenen, & B. Ruyter (Eds.), *EUSAI 2003: Proceeding of Ambient Intelligence, First European Symposium* (Vol. 2875, p. 15-26). Veldhoven, The Netherlands: Springer LNCS.

Krüger, A., & Kruppa, M. (2006). Integration of Mobile and Stationary Presentation Devices. In O. Stock & M. Zancanaro (Eds.), *PEACH: Intelligent Interfaces for Museum Visits.* Berlin, Heidelberg, New York: Springer Cognitive Technologies series, to appear.

Krüger, A., Kruppa, M., Müller, C., & Wasinger, R. (2002). Readapting Multimodal Presentations to Heterogenous User Groups. In R. Malaka & A. Krüger (Eds.), *ISAMP 2002: Notes of the AAAI-Workshop on Intelligent and Situation-Aware Media and Presentations (Technical Report WS-02-08, p. 46-54).* Edmonton, Canada: AAAI Press.

Kruppa, M. (2001). *ARREAL- A Mobile Pedestrian Navigation System Built From Standard Hardware Components.* (Diploma Thesis, Lehrstuhl Wahlster, Saarland University)

Kruppa, M. (2004). The Better Remote Control: Multiuser Interaction With Public Displays. In A. Butz, C. Kray, A. Krüger, & A. Schmidt (Eds.), *MU3I 2004: Proceedings of the First Workshop on Multi-User and Ubiquitous User Interfaces held in conjunction with IUI 2004* (p. 1-5). Funchal, Portugal: Sonderforschungsbereich 378 - Ressourcenadaptive kognitive Prozesse (Memo 83).

Kruppa, M. (2005). Towards Explicit Physical Object Referencing. In L. Ardissono, P. Brna, & A. Mitrovic (Eds.), *UM 2005: Proceedings of the Tenth User Modeling Conference 2005, Doctoral Consortium* (Vol. 3538, p. 506-508). Edinburgh, Scotland: Springer LNCS.

Kruppa, M., & Aslan, I. (2005). Parallel Presentations for Heterogenous User Groups - An Initial User Study. In M. Maybury, O. Stock, & W. Wahlster (Eds.), *INTETAIN 2005: Proceedings of the First Conference on Intelligent Technologies for Interactive Entertainment* (Vol. 3814, p. 54-63). Madonna de Campiglio, Italy: Springer LNCS.

Kruppa, M., Heckmann, D., & Krüger, A. (2005). Adaptive Multimodal Presentation of Multimedia Content in Museum Scenarios. *KI - Zeitschrift für Künstliche Intelligenz, 1*(5), 56-59.

Kruppa, M., & Krüger, A. (2003). Concepts for a Combined use of Personal Digital Assistants and Large Remote Displays. In T. Schulze, S. Schlechtweg, & V. Hinz (Eds.), *SimVis 2003: Proceedings of Simulation und Visualisierung* (p. 349-361). Magdeburg, Germany: SCS Publishing House e.V.

Kruppa, M., & Krüger, A. (2005). Performing Physical Object References With Migrating Virtual Characters. In M. Maybury, O. Stock, & W. Wahlster (Eds.), *INTETAIN 2005: Proceedings of the First Conference on Intelligent Technologies for Interactive Entertainment* (Vol. 3814, p. 64-73). Madonna de Campiglio, Italy: Springer LNCS.

Kruppa, M., Krüger, A., Rocchi, C., Stock, O., & Zancanaro, M. (2003). Seamless Personalized TV-like Presentations on Mobile and Stationary Devices in a Museum. In *ICHIM 2003: Proceedings of the International Conference on Hypermedia and Interactivity in Museums (CD-ROM).* Paris, France: Archives and Museum Informatics.

Kruppa, M., Spassova, M., & Schmitz, M. (2005). The Virtual Room Inhabitant - Intuitive Interaction With Intelligent Environments. In S. Zhang & R. Jarvis (Eds.), *AI 2005: Proceedings of the Eighteenth Australian Joint Conference on Artificial Intelligence* (Vol. 3809, p. 225-234). Sydney, Australia: Springer LNCS.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The Persona Effect: Affective impact of animated pedagogical agents. In *CHI 1997: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (p. 359-366). Atlanta, USA: ACM Press.

Lester, J. C., Towns, S. G., Callaway, C. B., Voerman, J. L., & FitzGerald, P. J. (2000). Deictic and Emotive Communication in Animated Pedagogical Agents. *Embodied Conversational Agents*, 123-154.

Lester, J. C., Voerman, J. L., Towns, S. G., & Callaway, C. B. (1999). Deictic Believability: Coordinated Gesture, Locomotion, and Speech in Lifelike Pedagogical Agents. *Applied Artificial Intelligence*, *13*(4-5), 383-414.

Lester, J. C., Zettlemoyer, L., Grégoire, J., & Bares, W. (1999). Explanatory Lifelike Avatars: Performing User-Centered Tasks in 3D Learning Environments. In *AGENTS 1999: Proceedings of the Third International Conference on Autonomous agents* (p. 24-31). Seattle, United States: ACM Press.

Lohse, M. (2001). *Ein System zur Visualisierung von Navigationsauskünften auf stationären und mobilen Geräten.* (Diploma Thesis, Lehrstuhl Wahlster, Saarland University)

Lucena, P. (2002). *Expressive Talking Heads: Um Estudo de Fala e Expressão Facial em Personagens Virtuais.* (PhD Thesis, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brazil)

Lucena, P., Gattass, M., & Velho, L. (2002). Expressive Talking Heads: A Study on Speech and Facial Expression in Virtual Characters. In *WTDCGPI 2002: Proceedings of the Workshop of Theses and Dissertations held in conjunction with SIBGRAPI 2002* (p. 4-7). Fortaleza, Brazil: IEEE Computer Society Press.

Lynch, K. (1960). *The Image of the City.* Cambridge, USA: MIT Press.

Lyons, J. (1977). *Semantics* (Vol. 2). London, UK: Cambridge University Press.

Malaka, R. (visited 2005). *EML - DEEPMAP project homepage.* (http://www.eml-development.de/deutsch/forschung/deepmap.php)

Malaka, R., Haeussler, J., & Aras, H. (2004). SmartKom Mobile: Intelligent Ubiquitous User Interaction. In *IUI 2004: Proceedings of the Ninth International Conference on Intelligent User Interfaces* (p. 310-312). Funchal, Madeira, Portugal: ACM Press.

Marsella, S., Gratch, J., & Rickel, J. (2004). Expressive Behaviours for Virtual Worlds. In H. Prendinger & M. Ishizuka (Eds.), *Life-Like Characters - Tools, Affective Functions, and Applications* (p. 317-360). Berlin, Heidelberg, New York: Springer Cognitive Technologies.

Mase, K. (1997). Aspects of Interface Agents: Avatar, Assistant and Actor. In *Workshop on Animated Interface Agents: Making Them Intelligent, held in conjunction with IJCAI 1997* (p. 33-37). Nagoya, Japan.

Mase, K., Sumi, Y., & Kadobayashi, R. (2000). The Weaved Reality: What Context-aware Interface Agents Bring About. In *ACCV 2000: Proceedings of the Asian Conference on Computer Vision* (p. 1120-1124). Taipei, Taiwan.

May, M. (1992). *Mentale Modelle von Städten. Wissenspsychologische Untersuchungen am Beispiel der Stadt Münster*. Münster/New York: Waxmann.

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.* (Technical Report Dartmouth Colledge)

McCulloch, W., & Pitts, W. (1943). A Logical Calculus of the IdeasImmanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, *5*, 115-137.

McNeill, D. (1992). *Hand and Mind - What Gestures Reveal about Thought (Context).* Chicago, USA: University of Chicago.

Metz, C. (1974). *Film Language: A Semiotics of the Cinema.* New York, USA: Oxford University Press.

Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems*, *E77-D*(12), 1321-1329.

Montello, D. R. (1998). A new Framework for Understanding the Acquisition of Spatial Knowledge in Large-Scale Environments. In M. J. Eganhofer & R. Golledge (Eds.), *Spatial and Temporal Reasoning in Geographic Information Systems* (p. 143-154). New York, USA: Oxford University Press.

Mulken, S. V., André, E., & Müller, J. (1998). The Persona Effect: How Substantial Is It? In J. May, J. Siddiqi, & J. Wilkinson (Eds.), *HCI 1998: Proceedings of Human Computer Interaction on People and Computers* (p. 53-66). Sheffield, UK: Springer.

Müller-Tomfelde C., & Reischl W. (1998). Communication Chairs: Examples of Mobile Roomware. In *CHI 1998: Proceedings of the Conference on Human Factors in Computing Systems (Summary), Late Breaking Results: Suite: The Real and the Virtual: Integrating Architectural and Information Spaces* (Vol. 2, p. 267-268). Los Angeles, USA: ACM.

Myers, B. (2001). Using Hand-Held Devices and PCs Together. *Communications of the ACM*, *44*(11), 34-41.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can Computer Personalities be Human Personalities? *International International Journal of Human-Computer Studies*, *43*(2), 223-239.

Nissenbaum, H. (1994). Computing and accountability. *Communications of the ACM*, *37*(1), 72-80.

Oppermann, R., & Specht, M. (2000). A Context-Sensitive Nomadic Exhibition Guide. In P. J. Thomas & H.-W. Gellersen (Eds.), *HUC 2000: Proceedings of the Second International Symposium on Handheld and Ubiquitous Computing* (Vol. 1927, p. 127-142). Bristol, UK: Springer LNCS.

Oppermann, R., Specht, M., & Jaceniak, I. (1999). Hippie: A Nomadic Information System. In H.-W. Gellersen (Ed.), *HUC 1999: Proceedings of the First International Symposium on Handheld and Ubiquitous Computing* (Vol. 1707, p. 330-333). Karlsruhe, Germany: Springer LNCS.

Pascoe, J. (1998). Adding Generic Contextual Capabilities to Wearable Computers. In *Proceedings of the Second International Symposium on Wearable Computers* (p. 92-99). Los Alamitos, USA: IEEE Computer Society Press.

Perdikeas, M. K., Chatzipapadopoulos, F. G., Venieris, I. S., & Marino, G. (1999). Mobile Agent Standards and Available Platforms. *Computer Networks*, *31*(19), 1999-2016.

Petta, P., & Trappl, R. (1997). Creating Personalities for Synthetic Actors, Towards Autonomous Personality Agents. In R. Trappl & P. Petta (Eds.), *Creating Personalities for Synthetic Actors* (Vol. 1195). Springer LNCS.

Pfleger, N., & Alexandersson, J. (2004). Modeling Non-Verbal Behavior in Multimodal Conversational Systems. In W. Wahlster (Ed.), *Special Journal Issue "Conversational User Interfaces", IT - Information Technology* (Vol. 46, p. 341-345). Munich, Germany: Oldenbourg Wissenschaftsverlag.

Pham, T., Schneider, G., & Goose, S. (2000). A Situated Computing Framework for Mobile and Ubiquitous Multimedia Access Using Small Screen and Composite Devices. In *Multimedia 2000: Proceedings of the Eigth International Conference on Multimedia* (p. 323-331). Marina del Rey, USA: ACM.

Pinhanez, C. S. (2001). The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. In G. D. Abowd, B. Brumitt, & S. Shafer (Eds.), *UbiComp 2001: Proceedings of the Third International Conference on Ubiquitous Computing* (Vol. 2201, p. 315-331). Atlanta, USA: Springer LNCS.

Prendinger, H., Descamps, S., & Ishizuka, M. (2002). Scripting Affective Communication with Life-Like Characters in Web-Based Interaction Systems. *Applied Artificial Intelligence*, *16*(7-8), 519-553.

Prendinger, H., & Ishizuka, M. (2004). Introducing the Cast for Social Computing: Life-Like

Characters. In H. Prendinger & M. Ishizuka (Eds.), *Life-Like Characters - Tools, Affective Functions, and Applications* (p. 3-16). Berlin, Heidelberg, New York: Springer Cognitive Technologies.

Prendinger, H., Ma, C., Yingzi, J., Nakasone, A., & Ishizuka, M. (2005). Understanding the Effect of Life-Like Interface Agents Through Users' eye Movements. In *ICMI 2005: Proceedings of the Seventh International Conference on Multimodal Interfaces* (p. 108-115). Trento, Italy: ACM Press.

Pulkki, V. (1997). Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society*, *45*(6), 456-466.

Pullmann, P. (1996). *The Golden Compass.* Knopf Books for Young Readers; 1st edition.

Rae, R. (2000). *Gestikbasierte Mensch-Maschine-Kommunikation auf der Grundlage visueller Aufmerksamkeit und Adaptivität.* (PhD Thesis, Technische Fakultät, Universität Bielefeld, Bielefeld, Germany)

Rae, R., Fislage, M., & Ritter, H. (1999). Visuelle Aufmerksamkeitssteuerung zur Unterstützung gestikbasierter Mensch-Maschine Interaktion. *KI - Künstliche Intelligenz, Themenheft Aktive Sehsysteme*, *99*(1), 18-24.

Rae, R., & Ritter, H. (visited 2005). *Active Vision Research at the Neuroinformatics Group - Homepage.* (http://www.techfak.uni-bielefeld.de/ags/ni/projects/camhead/)

Raptis, D., Tselios, N., & Avouris, N. (2005). Context-Based Design of Mobile Applications for Museums: A Survey of Existing Practices. In M. Tscheligi, R. Bernhaupt, & K. Mihalic (Eds.), *Mobile HCI 2005: Proceedings of the Seventh International Conference on Human Computer Interaction With Mobile Devices and Services* (p. 153-160). Salzburg, Austria: ACM.

Raubal, M., & Worboys, M. (1999). A Formal Model of the Process of Wayfinding in Built Environments. *Lecture Notes in Computer Science*, *1661*, 381-399.

Reeves, B., & Nass, C. (1998). *The Media Equation: How People Treat Computers, Television, and new Media Like Real People and Places.* New York, USA: Cambridge University Press.

Rehm, M., & André, E. (2005). Catch Me If You Can - Exploring Lying Agents in Social Settings. In F. Dignum, V. Dignum, S. Koenig, S. Kraus, M. P. Singh, & M. Wooldridge (Eds.), *AAMAS 2005: Proceedings of the International Conference on Autonomous Agents and Multiagent Systems* (p. 937-944). Utrecht, The Netherlands: ACM.

Retz-Schmidt, G. (1988). Various Views on Spatial Prepositions. *AI Magazine*, *9*(2), 95–105.

Rickel, J., & Johnson, W. L. (1999). Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *Applied Artificial Intelligence*, *14*(4-5), 343-382.

Rist, T., André, E., Baldes, S., Gebhard, P., Klesen, M., Kipp, M., Rist, P., , & Schmitt, M. (2004). A Review of the Development of Embodied Presentation Agents and Their Application Fields. In H. Prendinger & M. Ishizuka (Eds.), *Life-Like Characters - Tools, Affective Functions, and Applications* (p. 377-404). Berlin, Heidelberg, Germany: Springer Cognitive Technologies series.

Rist, T., Baldes, S., Gebhard, P., Kipp, M., Klesen, M., Rist, P., & Schmitt, M. (2002). CrossTalk: An Interactive Installation with Animated Presentation Agents. In E. André, A. Clarke, C. Fencott, G. Mitchell, & F. Nack (Eds.), *COSIGN 2002: Proceedings of the Second Conference on Computational Semiotics for Games and New Media* (p. 61-67). Augsburg, Germany: University of Augsburg.

Rist, T., & Schmitt, M. (2002). Applying Socio-Psychological Concepts of Cognitive Consistency to Negotiation Dialog Scenarios With Embodied Conversational Characters. In R. Aylett & L. Cañamero (Eds.), *AISB'02: Proceedings of the Symposium on Animated Expressive Characters for Social Interactions* (p. 79-84). London, UK: Imperial College, London.

Rocchi, C., Stock, O., Zancanaro, M., Kruppa, M., & Krüger, A. (2004). The Museum Visit: Generating Seamless Personalized Presentations on Multiple Devices. In *IUI 2004: Proceedings of the Ninth International Conference on Intelligent User Interfaces* (p. 316-318). Funchal, Madeira, Portugal: ACM Press.

Rocchi, C., & Zancanaro, M. (2004). Template-Based Adaptive Video Documentaries. In J. Baus, C. Kray, & R. Porzel (Eds.), *Proceedings of AIMS2004 Workshop, held in conjunction with UbiComp 2004* (p. 79-83). Nottingham, UK: Sonderforschungsbereich 378 - Ressourcenadaptive kognitive Prozesse (Memo 84).

Rodden, T., Cheverst, K., Davies, N., & Dix, A. (1998). Exploiting Context in HCI Design for Mobile Systems. In C. Johnson (Ed.), *Proceedings of the First Workshop on Human Computer Interaction with Mobile Devices* (p. 12-17). Glasgow, UK: University of Glasgow (GIST Technical Report G98-1).

Russel, S., & Norvig, P. (2002). *Artificial Intelligence: A Modern Approach (2nd Edition).* New York: Prentice Hall.

Sadalla, E., Burroughs, W., & Staplin, L. (2001). Reference Points in Spatial Cognition. *Journal of Experimental Psychology: Human Learning and Memory*(6), 516-528.

Schilit, B. N., Adams, N. I., & Want, R. (1994). Context-Aware Computing Applications. In *Proceedings of the Workshop on Mobile Computing Systems and Applications* (p. 85-90). Los Alamitos, USA: IEEE Computer Society Press.

Schilit, B. N., & Theimer, M. (1994). Disseminating Active Map Information to Mobile Hosts. *IEEE Network*, *8*(5), 22-32.

Schirra, J. R. J. (1994). *Bildbeschreibung als Verbindung von visuellem und sprachlichem Raum. Eine interdisziplinäre Untersuchung von Bildvorstellungen in einem Hörermodell.* Amsterdam, The Netherlands: PhD. thesis, Dissertationen zur Künstlichen Intelligenz, Vol. 71, IOS Press.

Schmitt, M. (2005). *Dynamische modellierung interpersoneller Beziehungen zwischen virtuellen Charakteren.* Saarbrücken, Germany: Universität des Saarlandes.

Schmitz, M. (2004). SAFIR: Spatial Audio Framework for Instrumented Rooms. In A. Quigley & S. Tanimoto (Eds.), *IT 2004: Workshop on Invisible and Transparent Interfaces held in conjunction with AVI 2004* (p. 44-50). Gallipoli, Italy.

Schuler, D. (1994a). Social computing. *Communications of the ACM*, *37*(1), 28-29.

Schuler, D. (1994b). Community Networks: Building a new Participatory Medium. *Commununications of the ACM*, *37*(1), 38-51.

Shoemaker, G. B. D., & Inkpen, K. M. (2001). Single Display Privacyware: Augmenting Public Displays With Private Information. In *CHI 2001: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (p. 522-529). New York, USA: ACM Press.

Spassova, L. (2004a). Fluid Beam - A Steerable Projector and Camera Unit. In *ISWC/ISMAR 2004: Proceedings of the Student and Newbie Colloquium* (p. 56-58). Arlington, USA: Electronic Laboratory, ETH Zürich.

Spassova, L. (2004b). *Fluid Beam - Konzeption und Realisierung eines Display-Kontinuums mittels einer steuerbaren Projektor-Kamera-Einheit.* (Diploma Thesis, Lehrstuhl Wahlster, Saarland University)

Spassova, L., Wasinger, R., Baus, J., & Krüger, A. (2005). Product Associated Displays in a Shopping Scenario. In *ISMAR 2005: Proceedings of the Fourth IEEE / ACM International Symposium on Mixed and Augmented Reality* (p. 210-211). Vienna, Austria: IEEE Computer Society.

Stahl, C., Baus, J., Krüger, A., Heckmann, D., Wasinger, R., & Schneider, M. (2004). REAL: Situated Dialogues in Instrumented Environments. In A. Quigley & S. Tanimoto (Eds.), *ITI 2004: Workshop on Invisible and Transparent Interfaces held in conjuction with AVI 2004* (p. 10-15). Gallipoli, Ítaly.

Stahl, C., Baus, J., Krüger, A., & Schmitz, M. (2005). Managing Presentations in an Intelligent Environment. In A. Butz, C. Kray, A. Krüger, A. Schmidt, & H. Prendinger (Eds.), *MU3I*

2005: Proceedings of the Second Workshop on Multi-User and Ubiquitous User Interfaces held in conjunction with IUI 2005 (p. 3-4). San Diego, U.S.A.: Sonderforschungsbereich 378 - Ressourcenadaptive kognitive Prozesse (Memo 85).

Stifelmann, L. J. (1994). *The Cocktail Party Effect in Auditory Interfaces: A Study of Simultaneous Presentation* (Tech. Rep.). Massachusetts, USA: MIT Media Laboratory.

Stocky, T. (2002). *Conveying Routes: Multimodal Generation and Spatial Intelligence in Embodied Conversational Agents.* (Master of Engineering Thesis, Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, USA)

Stocky, T., & Cassell, J. (2002). Shared Reality: Spatial Intelligence in Intuitive User Interfaces. In *IUI 2002: Proceedings of the Seventh International Conference on Intelligent User Interfaces* (p. 224-225). San Francisco, USA: ACM Press.

Sumi, Y., Etani, T., Fels, S., Simonet, N., Kobayashi, K., & Mase, K. (1998). C-MAP: Building a Context-Aware Mobile Assistant for Exhibition Tours. In T. Ishida (Ed.), *Community Computing and Support Systems, Social Interaction in Networked Communities* (Vol. 1519, p. 137-154). London, UK: Springer LNCS.

Sumi, Y., & Mase, K. (2001a). AgentSalon: Facilitating Face-to-Face Knowledge Exchange Through Conversations Among Personal Agents. In *Agents 2001: Proceedings of the Fifth International Conference on Autonomous Agents* (p. 393-400). Montreal, Canada: ACM Press.

Sumi, Y., & Mase, K. (2001b). AgentSalon: Supporting new Encounters and Knowledge Exchanges by Chats of Personal Agents. In *CHI 2001: Extended Abstracts on Human factors in Computing Systems* (p. 191-192). Seattle, Washington: ACM Press.

Sumi, Y., & Mase, K. (2004). Interface Agents That Facilitate Knowledge Interactions Between Community Members. In H. Prendinger & M. Ishizuka (Eds.), *Life-Like Characters - Tools, Affective Functions, and Applications* (p. 405-427). Berlin, Heidelberg, New York: Springer Cognitive Technologies series.

Suwa, M., Sugie, N., & Fujimora, K. (1978). A Preliminary Note on Pattern Recognition of Human Emotional Expression. In *Proceedings of the Fourth International Joint Conference on Pattern Recognition* (p. 408-410). Kyoto, Japan: IEEE.

Terrenghi, L. (2005). Design for Interaction in Instrumented Environments. In H. W. Gellersen, R. Want, & A. Schmidt (Eds.), *Advances in Pervasive Computing: Adjunct Proceedings of the Third International Conference on Pervasive Computing* (Vol. 3468, p. 129-136). Munich, Germany: Springer LNCS.

Thalmann, D., Noser, H., & Huang, Z. (1997). Autonomous Virtual Actors Based on Virtual

Sensors. In P. Petta & R. Trappl (Eds.), *Creating Personalities for Synthetic Actors* (Vol. 1195, p. 25-42). Berlin, Germany: Springer LNCS.

Tian, Y.-L., Kanade, T., & Cohn, J. (2003). Facial expression analysis. In S. Li & A. Jain (Eds.), *Handbook of face recognition.* Berlin, Heidelberg, New York: Springer.

Towns, S. G., Callaway, C. B., & Lester, J. C. (1998). Generating Coordinated Natural Language and 3D Animations for Complex Spatial Explanations. In *AAAI 1998/IAAI 1998: Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence* (p. 112-119). Menlo Park, USA: American Association for Artificial Intelligence.

Towns, S. G., Voerman, J. L., Callaway, C. B., & Lester, J. C. (1998). Coherent Gestures, Locomotion, and Speech in Life-Like Pedagogical Agents. In *IUI 1998: Proceedings of the Third International Conference on Intelligent User Interfaces* (p. 13-20). San Francisco, USA: ACM Press.

Trimble Navigation Limited. (visited 2005). *Gps-tutorial.* (Available at: http://www.trimble.com/gps/index.html)

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *49*, 433-460.

Tversky, B., & Lee, P. U. (1999). Pictorial and Verbal Tools for Conveying Routes. In C. Freksa & D. M. Mark (Eds.), *COSIT 1999: Proceedings of the International Conference on Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science* (Vol. 1661, p. 51-64). Stade, Germany: Springer LNCS.

Wahlster, W. (2003). Towards Symmetric Multimodality: Fusion and Fission of Speech, Gesture and Facial Expression. In A. Günter, R. Kruse, & B. Neumann (Eds.), *KI 2003: Advances in Artificial Intelligence, Proceedings of the Twentysixth German Conference on Artificial Intelligence* (Vol. 2821, p. 1-18). Hamburg, Germany: Springer LNCS.

Wahlster, W. (2005). Overview of the Current State of the Lead Innovation VirtualHuman. In B. Reuse, W. Wahlster, & J. Encarna (Eds.), *VirtualHuman: Slides of the 3rd Project Meeting and VirtualHuman Network of Excellence Meeting held in conjunction with INTETAIN 2005.* Madonna de Campiglio, Italy: available online at http://www.virtual-human.org/.

Wahlster, W., Baus, J., Kray, C., & Krüger, A. (2001). REAL: Ein ressourcenadaptierendes mobiles Navigationssystem. *Informatik Forschung und Entwicklung*, *16*(4), 233-241.

Wahlster, W., & Kobsa, A. (1989). User Models in Dialog Systems. In A. Kobsa & W. Wahlster (Eds.), *User Models in Dialog Systems.* Berlin, Germany: Springer (Symbolic Computation Series).

Wahlster, W., Reithinger, N., & Blocher, A. (2001a). SmartKom: Multimodal Communication with a Life-Like Character. In *EUROSPEECH 2001: Proceedings of the Seventh European Conference on Speech Communication and Technology* (Vol. 3, p. 1547-1550). Aalborg, Denmark: ISCA.

Wahlster, W., Reithinger, N., & Blocher, A. (2001b). SmartKom: Towards Multimodal Dialogues with Anthropomorphic Interface Agents. In G. Wolf & G. Klein (Eds.), *Proceedings of International Status Conference - Human-Computer Interaction* (p. 23-34). Saarbrücken, Germany: Deutsches Zentrum für Luft- und Raumfahrt (DLR) e.V. (German Aerospace Center) on the behalf of the BMWF, Referat 524 (Hg.).

Wahlster, W., & Tack, W. (1998). SFB 378: Ressourcenadaptive Kognitive Prozesse. SFB 378 (REAL). In M. Jarke, K. Pasedach, & K. Pohl (Eds.), *Informatik 1997: Informatik als Innovationsmotor, 27. Jahrestagung der Gesellschaft für Informatik* (p. 51-57). Aachen, Germany: Springer.

Want, R., Adams, N. I., Schilit, B. N., Gold, R., Petersen, K., Ellis, J. R., Goldberg, D., & Weiser, M. (1995). An Overview of the PARCTAB Ubiquitous Computing Experiment. *IEEE Personal Communications*, *2*(6), 28-34.

Want, R., & Hopper, A. (1992). Active Badges and Personal Interactive Computing Objects. *IEEE Transactions on Consumer Electronics*, *38*(1), 10-20.

Want, R., Hopper, A., Falcao, V., & Gibbons, J. (1992). The Active Badge Location System. *ACM Transactions on Information Systems*, *10*(1), 91-102.

Want, R., Schilit, B. N., Adams, N. I., Gold, R., Petersen, K., Goldberg, D., Ellis, J. R., & Weiser, M. (1995). *The PARCTAB Ubiquitous Computing Experiment* (Tech. Rep. No. CSL-95-1). Palo Alto, USA: Xerox Palo Alto Research Center.

Ward, A., Jones, A., & Hopper, A. (1997). A New Location Technique for the Active Office. *IEEE Personal Communications*, *4*(5), 42-47.

Wasinger, R., Krüger, A., & Jacobs, O. (2005). Integrating Intra and Extra Gestures Into a Mobile and Multimodal Shopping Assistant. In H.-W. Gellersen, R. Want, & A. Schmidt (Eds.), *Pervasive: Proceedings of the Third International Conference on Pervasive Computing* (Vol. 3468, p. 297-314). Munich, Germany: Springer LNCS.

Wasinger, R., & Wahlster, W. (2006). The Anthropomorphized Product Shelf: Symmetric Multimodal Interaction with Instrumented Environments. In E. Aarts & J. Encarnação (Eds.), *True Visions: The Emergence of Ambient Intelligence.* Heidelberg, Germany: Springer.

Weiser, M. (1991). The Computer for the 21st Century. *Scientific American Magazine*, *265*(3), 94-104.

Weizenbaum, J. (1966). ELIZA - A computer Program for the Study of Natural Language Communication Between Man and Machine. *Commununications of the ACM*, *9*(1), 36-45.

Yantis, S., & Jonides, J. (1990). Abrupt Visual Onsets and Selective Attention: Voluntary Versus Automatic Allocation. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 121-134.

Yerazunis, W., & Carbone, M. (2001). *Privacy-Enhanced Displays by Time-Masking Images* (Tech. Rep. No. TR2002-011). Cambridge, USA: Mitsubishi Electronic Research Laboratories.

Index

N

O

P