Tangible Interaction with Anthropomorphic Smart Objects in Instrumented Environments

Dissertation
zur Erlangung des Grades
des Doktors der Ingenieurwissenschaften (Dr.-Ing.)
der Naturwissenschaftlich-Technischen Fakultäten der Universität des Saarlandes

vorgelegt von
Michael Schmitz

Saarbrücken
25. November 2010

Dekan:
Prof. Dr. Holger Hermanns
Vorsitzender des Prüfungsausschusses:
Prof. Dr. Philipp Slusallek
Berichterstatter:
Prof. Dr. Dr. h.c. mult. Wolfgang Wahlster
Prof. Dr. Antonio Krüger
Promovierter akademischer Mitarbeiter der Fakultät:
Dr. Dominikus Heckmann
Tag des Kolloquiums:
22.12.2010

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbständig 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.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form in einem Verfahren zur Erlangung eines akademischen Grades vorgelegt.

Michael Schmitz, Saarbrücken, 26. November 2010

Acknowledgments

First of all, I would like to thank Prof. Wahlster for the opportunity to work on this dissertation and for the extensive and productive supervision at the very final stage of this work. I am further much obliged to Prof. Krüger for taking the role of the second advisor and to my colleagues (in alphabetical order) Jörg Baus, Doris Borchers, Boris Brandherm, Benedict Fehringer, Patrick Gebhard, Dominikus Heckmann, Ralf Jung, Gerrit Kahl, Lübomira Spassova (and probably a few more) for proofreading, advise, discussions, ideas, and general support.

I also have to express my gratitude to all students, who contributed a significant amount of work, particularly Yanchuan Li, Sarah Schmidt, Robert Dörr, Moritz Federlein, Henning Zimmer, Ying Wang, and Benjamin Hollinger. They all shared their distinct and unique abilities to add important bits and pieces to the result of this dissertation.

Last not least, I have to thank my family and my friends for explicit and implicit, requested and unexpected support of all kinds. I suppose, I should mention Nico Czaja in particular, this extraordinary, intellectual hungarian writer and friend, who did some in-depth proofreading of several parts of this thesis (including these acknowledgements). I still owe him a tiny, but shiny surprise present for that.

And, im Grunde genommen, thank you, Anusche.

Kurzzusammenfassung

Die Integration von Informationstechnologie in Gebrauchsgegenstände ist ein gegenwärtiger technologischer Trend, welcher es Alltagsgegenständen ermöglicht, durch den Einsatz von Sensorik, Aktorik und drahtloser Kommunikation neue Dienste anzubieten, die über den ursprünglichen Zweck des Objekts hinausgehen. Die Nutzung dieser sogenannten Smart Objects erfordert neuartige Benutzerschnittstellen, welche die speziellen Eigenschaften und Anwendungsbereiche solcher Systeme berücksichtigen. Konzepte aus den Bereichen Tangible Interaction und Anthropomorphe Benutzerschnittstellen werden in dieser Dissertation vereint, um ein neues Interaktionsparadigma für Smart Objects zu entwickeln. Die vorliegende Arbeit untersucht dafür die Gestaltungsmöglichkeiten und zeigt relevante Aspekte aus verwandten Disziplinen auf. Darauf aufbauend werden Richtlinien eingeführt, welche den Entwurf von Benutzerschnittstellen nach dem hier vorgestellten Ansatz begleiten und unterstützen sollen. Für eine prototypische Implementierung solcher Benutzerschnittstellen wird eine Architektur vorgestellt, welche die Anforderungen von Smart Object Systemen in instrumentierten Umgebungen berücksichtigt. Ein wichtiger Bestandteil stellt dabei die Sensorverarbeitung dar, welche unter anderem eine Interaktionserkennung am Objekt und damit auch eine physikalische Eingabe ermöglicht. Des Weiteren werden neuartige Methoden für den auditiven Ausdruck von Emotion und Persönlichkeit entwickelt, welche essentielle Bausteine für anthropomorphisierte Smart Objects darstellen und in Benutzerstudien untersucht wurden. Die Dissertation schliesst mit der Beschreibung von drei Applikationen, welche im Rahmen der Arbeit entwickelt wurden und das Potential der hier erarbeiteten Konzepte und Methoden widerspiegeln.

Short Abstract

A major technological trend is to augment everyday objects with sensing, computing and actuation power in order to provide new services beyond the objects' traditional purpose, indicating that such smart objects might become an integral part of our daily lives. To be able to interact with smart object systems, users will obviously need appropriate interfaces that regard their distinctive characteristics. Concepts of tangible and anthropomorphic user interfaces are combined in this dissertation to create a novel paradigm for smart object interaction. This work provides an exploration of the design space, introduces design guidelines, and provides a prototyping framework to support the realisation of the proposed interface paradigm. Furthermore, novel methods for expressing personality and emotion by auditory means are introduced and elaborated, constituting essential building blocks for anthropomorphised smart objects. Two experimental user studies are presented, confirming the endeavours to reflect personality attributes through prosody-modelled synthetic speech and to express emotional states through synthesised affect bursts. The dissertation concludes with three example applications, demonstrating the potentials of the concepts and methodologies elaborated in this thesis.

Zusammenfassung

Der fortlaufende Trend der Informationstechnologie, immer kleinere, leistungsfähigere und billigere Prozessoren, Kommunikationseinheiten, Sensoren und Aktoren zu entwickeln, wird dazu führen, dass Computer wesentlich intensiver in unseren Alltag Einzug halten werden, als es bereits der Fall ist. Einzelne Objekte und ganze Umgebungen können so mit miniaturisierter Elektronik instrumentiert und dadurch mit zusätzlichen Fähigkeiten versehen werden, welche über die ursprüngliche Anwendung hinaus gehen. In dieser Arbeit werden sogenannte Smart Objects betrachtet, welche alltägliche Gegenstände sind, die durch die Integration von Informationstechnologien zusätzliche Dienste anbieten können. Um mit Smart Objects umgehen zu können, werden neuartige Benutzerschnittstellen benötigt, welche die speziellen Eigenschaften und Anwendungsbereiche solcher Systeme berücksichtigen.

Anthropomorphe Benutzerschnittstellen zielen darauf ab, bestehende Fähigkeiten des Menschen auszunutzen, welche in alltäglichen, sozialen Begegnungen erlernt werden. Die Disziplin der Tangible User Interfaces folgt ebenfalls dem Prinzip, auf vorhandenen Fähigkeiten des Menschen aufzubauen, konzentriert sich dabei aber unter anderem auf die motorischen Fähigkeiten zur physikalischen Manipulation von Objekten. Diese beiden Ansätze werden in dieser Dissertation vereint, um ein neues Interaktionsparadigma für Smart Objects zu entwickeln. Die vorliegende Arbeit untersucht dafür die Gestaltungsmöglichkeiten und zeigt relevante Aspekte aus verwandten Disziplinen auf. Darauf aufbauend werden Richtlinien eingeführt, welche den Entwurf von Benutzerschnittstellen nach dem hier vorgestellten Ansatz begleiten und unterstützen sollen.

Für eine prototypische Implementierung solcher Benutzerschnittstellen wird eine Architektur vorgestellt, die sich durch einen flexiblen, modularen Aufbau auszeichnet und die Anforderungen von Smart Object Systemen in instrumentierten Umgebungen berücksichtigt. Ein wichtiger Bestandteil stellt dabei die Sensorverarbeitung dar, welche unter anderem eine Interaktionserkennung am Objekt und damit auch die physikalische Eingabe ermöglicht. Dazu wurde eine Bibliothek entwickelt, welche eine Implementierung von Systemen auf Basis von Mikrocontroller-Plattformen vereinfacht, indem die Verarbeitung von Sensordaten in Form von Signalverarbeitungsketten strukturiert wird. Die Bibliothek beinhaltet eine Sammlung vorgefertigter Komponenten zur Anbindung von Mikrocontroller-Plattformen, Filterkomponenten zur Datenverarbeitung, sowie Ausgabekomponenten, welche die Ergeb-

nisse der Sensorverarbeitung beispielsweise an andere Softwaremodule weitergeben, auf dem Bildschirm anzeigen oder lokal speichern.

In vielen Fällen sind Smart Objects keine isolierten Systeme sondern kommunizieren mit instrumentierten Umgebungen oder sind sogar für die Ausführung von Funktionen auf sie angewiesen. Solche Umgebungen können nun beispielsweise Dienste anbieten, welche von Smart Objects bzw. den darauf aufbauenden Anwendungen genutzt werden können. Im Rahmen dieser Arbeit wurden drei solcher Dienste entwickelt: 1. Eine kamerabasierte Anwendung zur Positionsbestimmung von Personen in Räumen, um sensorisch erkannte Interaktionen zu Personen zuordnen zu können 2. Ein System für die Realisierung von räumlichem Audio, mit dem es Smart Objects ermöglicht wird, auch ohne eigene akustische Ausgabemöglichkeit den auditiven Kanal zu nutzen 3. Eine zentraler Dienst zur Verwaltung und Steuerung von Ausgabegeräten wie beispielsweise Bildschirmen, um den parallelen Einsatz mehrerer Anwendungen zu ermöglichen.

Für die Darstellung anthropomorpher Aspekte wurden zudem neue Verfahren entwickelt und in der vorliegenden Dissertation vorgestellt. Um die Ausprägung von bestimmten Persönlichkeitsdimensionen mit einem Text-To-Speech-System wiedergeben zu können, wurde ein Modell für den Einsatz prosodischer Parameter entworfen und evaluiert. Ergänzend dazu wird gezeigt, wie Dialogverhalten und Formulierungsalternativen dazu beitragen können, Persönlichkeitszügen des Sprechers darzustellen. Somit können sprechende Objekte zur Laufzeit passend zur zugeordneten Persönlichkeit aus prosodischen Parametersets und Formulierungsalternativen wählen. Außerdem wird in der Arbeit dargestellt, wie synthetische Affect Bursts aus unterschiedlichen Basistönen generiert werden können, um damit affektive Zustände zu kommunizieren, welche in ihrer Klangcharakteristik an das Objekt angepasst sind. Eine Benutzerstudie bestätigt die Erkennung der mit diesem Verfahren abgebildeten Emotionen.

Abschließend werden drei beispielhafte Anwendungen vorgestellt, welche das Potential der in dieser Arbeit erarbeiteten Konzepte und Methoden widerspiegeln. Eine instrumentierte Pillendose prüft die Einhaltung vorgegebener Lagerbedingungen und Einnahmen des Patienten und warnt bei Verstößen gegen Vorgaben beziehungsweise bestätigt dem Patienten korrektes Verhalten durch anthropomorphe akustische Signale. Beim digitalen Sommelier stellen sich Weinflaschen aus unterschiedlichen Regionen mit entsprechend unterschiedlichen Persönlichkeiten vor, wenn sie aus dem Einkaufsregal genommen werden. Der anthropomorphe Cocktailshaker gibt Anweisungen zum korrekten Umgang mit ihm selbst und reagiert auf zahlreiche physikalische Aktionen des Benutzers.

Contents

1 INTRODUCTION

Doolittle: *"Bomb, this is Doolittle. You are not to detonate, repeat, you are NOT to detonate in the bomb bay. Disarm yourself. This is an order."*
Bomb 20: *"I read you, Lieutenant Doolittle, but I am programmed to detonate in fourteen minutes. Detonation will occur at the programmed time."*
(Lieutenant Doolittle ponders about a way of preventing the detonation of bomb 20)
Doolittle: *"Fine. Think about this one, then: how do you know you exist?"*
Bomb 20: *"Well, of course I exist."*
Doolittle: *"But how do you know you exist?"*
Bomb 20: *"It is intuitively obvious."*
Doolittle: *"Intuition is no proof. What concrete evidence do you have of your own existence?"*
Bomb 20: *"Hmm... Well, I think, therefore I am."*
Doolittle: *"That's good. Very good. Now then, how do you know that anything else exists?"*
Bomb 20: *"My sensory apparatus reveals it to me."*
Doolittle: *"Right!"*
Bomb 20: *"This is fun."*
Doolittle: *"All right now, here's the big question: how do you know that the evidence your sensory apparatus reveals to you is correct? What I'm getting at is this: the only experience that is directly available to you is your sensory data. And this data is merely a stream of electrical impulses which stimulate your computing centre."*
Bomb 20: *"In other words, all I really know about the outside universe relayed to me through my electrical connections."*
Doolittle: *"Exactly!"*
Bomb 20: *"Why, that would mean... I really don't know what the outside universe is like at all, for certain."*
Doolittle: *"That's it!"*
Bomb 20: *"Intriguing. I wish I had more time to discuss this matter."*
Doolittle: *"Why don't you have more time?"*
Bomb 20: *"Because I must detonate in seventy-five seconds."*

(excerpts from a dialogue in John Carpenter's science fiction move *"Dark Star"*, 1974)

The preceding dialogue excerpt of the classic science fiction satire involves an anthropomorphic bomb that is equipped with a supposedly intuitive speech interface and ultimate autonomous behaviour. Whereas failure-proof speech interfaces are a common interface vision in science fiction movies and novels, the cited clip ironically demonstrates the absurdity of such an anthropomorphic interface for expert users in a highly critical situation. Not only that this form of a spoken dialogue constitutes a comparably slow interface in comparison to traditional command line input, the use of push-buttons or also direct manipulation techniques, the interactive object reduces the level of control of the user - which actually became a major argument against agent-based interfaces within the research community of human computer interaction. Although the talking bomb discloses several attributes of a typical counter-example for appropriate anthropomorphism, the idea of smart and interactive everyday objects (considering bombs as everyday objects as well) is a highly active research topic, indicating that smart objects might become an integral part of our daily lives. One technological trend is to augment such everyday objects with sensing, computing and actuation power in order to provide new services beyond the objects' traditional purpose. Sensor technology does not only facilitate new qualities of objects but also constitute the interactive layer between system and users as providing channels for human-computer interaction with physical devices - just as bomb 20 phrased it: ”My sensory apparatus reveals it to me”.

To be able to interact with smart object systems users will obviously need novel and appropriate interfaces, which hold for multiple domains and do not require an intensive learning phase, considering that a user might often be exposed to smart objects that demand spontaneous usage of unfamiliar systems. Such situations will require that a user should be able to build up a mental model quickly, which particularly holds for novice users.

People often tend to treat objects similar to humans, according to findings of Reeves and Nass [Reeves and Nass, 1996], which allows users to explain the behaviour of a system if they lack a good functional conceptual model. This is an important issue for the envisioned application scenarios, where users have to spontaneously deal with unfamiliar smart object environments. Other researchers conclude that for intuitive and efficient human-environment interaction with thousands of networked smart objects a ”limited animistic design metaphor” might be appropriate [Nijholt et al., 2004]. Similarly, Mithen [Mithen, 1996] pointed out that ”the most powerful metaphors are those which cross domain boundaries, such as by associating a living entity with something that is inert or an idea with something that is tangible.”

Anthropomorphic interfaces generally attempt to build on established human skills, learned in daily social encounters. The research domain of tangible user interfaces shares this goal but focuses on motor skills, mainly related to physical manipulation of tangible objects. Both concepts complement each other, and thus, we attempt to exploit the naturalness of conversational speech and the physicality of real world objects as the foundation for a new interaction paradigm for smart objects. We believe that the anthropomorphic approach in combination with a tangible interface will provide an intuitive user interface for quick and casual interaction in scenarios where learning time is not given or not appropriate. We consider the elderly and children as notably interesting user groups that have the potential to profit from such an anthropomorphic interface. It is further important to identify the limitations of this approach: For instance, we do not expect to be able to handle complex and recurring tasks as efficiently as specialised user interfaces tailored for a specific application. The main question of this exploratory work is, how we can design **T**angible interfaces for **A**nthropomorphic **S**mart **O**bjects (TASOs), considering potential application domains and general design limitations.

1.1 Aims and Methods

The objectives of this dissertation are of very interdisciplinary nature and require to approach the topic from a broader perspective, mainly drawing on human-computer interaction, artificial intelligence, cognitive and social psychology and ubiquitous computing. Within the research discipline of human-computer interaction this dissertation essentially combines three different subareas (as sketched in Figure 1.1).

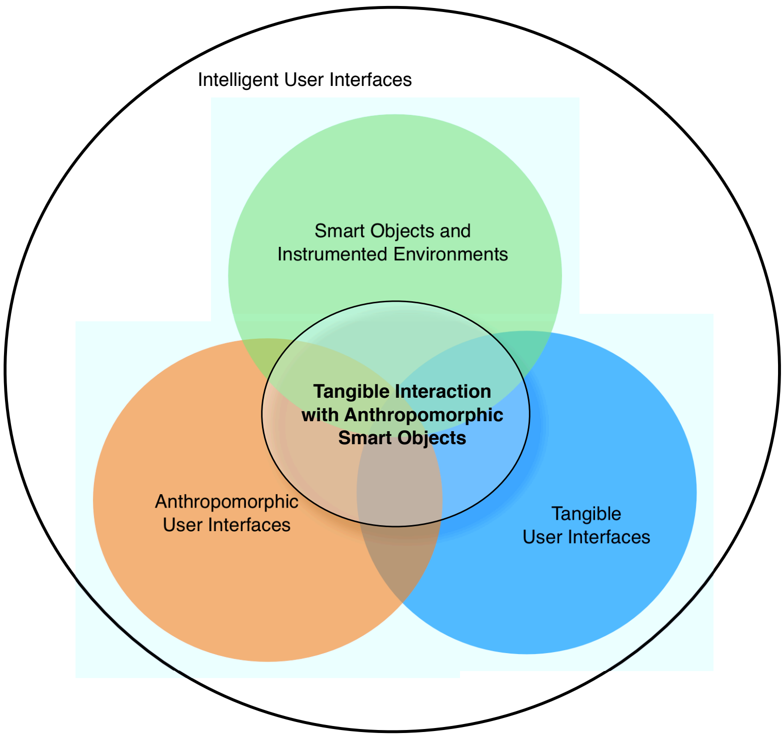


***Figure 1.1**: Combined research subareas*

The aims of this work are manyfold and include the aggregation and combination of related research in order to formulate design guidelines for the novel interaction paradigm for TASOs. The evolution of interfaces in the era of ubiquitous computing and related technologies has to be reflected for an analysis of the requirements posed on the design and realisation of smart object interaction in general. Technological considerations and their impact on the design process are summarised in Section 4.2 and applied in the development of the prototyping testbed introduced in Chapter 5.

At the moment, smart object solutions and their purpose-built interaction modes are rather proprietary and specialised. Considering the penetration of everyday life with smart object systems in a near future, a situation comprising an abundance of isolated solutions cannot be desirable. The aim of this work is to explore the possibilities and potential of the proposed paradigm towards a natural and intuitive smart object interaction. Findings of this work could have a major impact on smart object interaction design and influence future research in this field.

As a result of this work, we will develop guidelines for tangible, anthropomorphic interaction design for smart object interfaces, outlining application possibilities and guide interaction designers to create appropriate interfaces for the intended tasks. Rules will be identified for building interfaces that are intuitive to use and enjoyable for presumably casual and uncritical scenarios. The approach taken by this thesis requires the exploration of new interaction concepts based on speech and non-verbal sound, which will be realised and evaluated.

From the above stated considerations, we derive several research questions that we have to address in this dissertation:

1. **What are the desired attributes of TASOs?**

   During the course of this work we will define and refine the requirements and limitations of the proposed interaction paradigm and formulate guidelines and aims to support designers and developers in the process of creating instances of this paradigm.

2. **Which concepts of tangible interaction and anthropomorphism can be combined to design interfaces for smart objects and how can this entanglement generally be accomplished?**

   With an extensive review and analysis of related work we have to identify relevant contributions of previous work in the corresponding fields and derive an appropriate subset for the combined application of these paradigms. Since the total is expected to be more than the sum of its parts, novel elements have to be further explored during the course of this work.

3. **What is the design space of a TASO?**

   The design space of the proposed interface paradigm crosses a variety of domains. For a comprehensive account of the design dimensions involved in the creation of such interfaces we need to look outside the box and shed light on a number of issues that relate to the design process. We further have to analyse the resulting implications and include them into the proposed prototyping process.

4. **In which kind of interfaces can TASOs be applied?**

   Are TASOs a general interface for certain types of applications, or can we distinguish different categories of TASOs that can be applied for certain tasks and applications?

During the course of our research efforts we have to categorise appropriate use cases and identify appropriate classes of TASOs.

5. **What are the technological requirements for the realisation of TASOs?**

   Smart objects and instrumented environments exist in various forms and combinations, which pose several restrictions not only on the realisation but also on the overall design process of TASOs. We have to analyse and discuss technical issues and their impact to support a structured prototyping process that has to cope with a multitude of configurations.

6. **How can TASOs be efficiently and effectively prototyped?**

   In order to realise and explore instantiations of the TASO concept, we have to develop an efficient prototyping testbed that allows for flexible implementation and configuration of smart object interfaces in instrumented environments in a modular fashion that ensures a high reuse of developed modules. One important part will be a framework for sensor-interpretation and fusion as the foundation of object centric interaction recognition and actuation management.

7. **How can personality and emotions be effectively expressed in TASOs?**

   A major aspect of the TASO approach is the creation of consistent anthropomorphic entities in many variations. Established means of representing personality and emotion can be drawn upon, but novel methods in the domain of smart object interaction have to be developed and evaluated.

1.2 Chapter Outline

The general structure of this thesis is intended to provide several entry points to start reading and to be able to browse through this work, while concentrating on aspects that are of particular interest for the reader. Each of the following chapters starts with a brief introduction and an overview of its contents, and concludes with a summarising discussion reviewing the main points covered by this chapter. Figure 1.2 displays the structure of the dissertation in sequential order, visualising the categories, which the chapters have been grouped into.

The first part builds up the theoretical foundation for the remainder of this work and introduces related research and discusses how the chosen approach of combining tangible interaction, anthropomorphism and smart objects embeds into the research landscape and discloses new territories. This part starts with basic concepts of each research discipline, including important working definitions and an analysis of how the concepts of each research area complement each other and the potential benefits that arise from a combination. The third chapter reviews prominent related projects in each corresponding research area and also first attempts of combining them. This will be concluded by a discussion of open issues and typical problems that will be addressed by the TASO-approach. Chapter 4 describes the

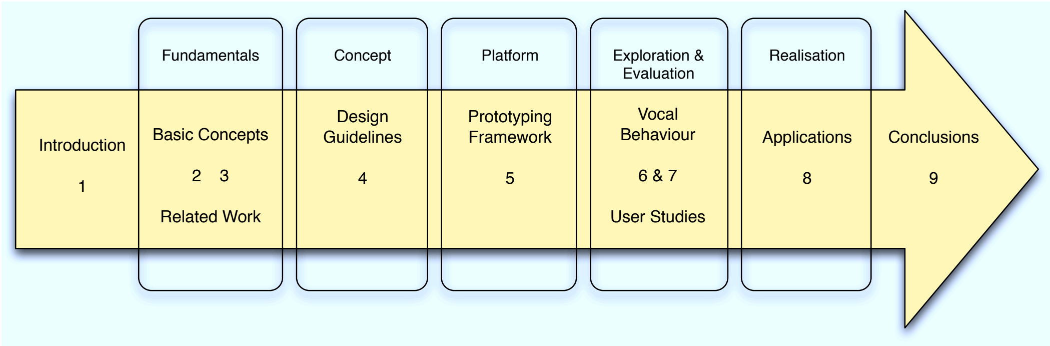


***Figure 1.2**: Structural overview*

conceptual part of this work and provides a discussion of the design space and guidelines for the development of TASO. This includes a categorisation of anthropomorphic objects for different use cases and a discussion of technical and conceptual constraints. The platform part starts off with the presentation of the prototyping testbed and its architecture in Chapter 5. This includes a programming library and a graphical toolkit for sensor-based development to support the realisation of anthropomorphic agents embodied by smart objects. The sensory layer will deal with the integration of distributed instrumentation that could reside either inside the smart object or also in the environment. The goal is to allow smart objects to detect their state and user interactions related to themselves, for instance object manipulation such as picking up or turning objects. This is followed by the part covering exploration and evaluation, where we will elaborate (Chapter 6) and evaluate (Chapter 7) novel means of generating vocal behaviour expressing emotion and personality in acoustic displays TASO. Chapter 8 describes realisations of concepts that have been introduced in Chapter 4, building on the technological infrastructure specified in Chapter 5 and on the expressive means introduced in Chapter 6. This includes the Sensing Pill Box, the Digital Sommelier and the Anthropomorphic Cocktail Shaker. The dissertation concludes in Chapter 9 with a summary of the scientific results and contributions, and a discussion of potential future work emerging from the outcomes of this thesis.

2 BACKGROUND & BASIC CONCEPTS

The aim of this chapter is to provide an adequate background for the subsequent parts of this work and to embed the topic into the current research context.

This chapter is organised in five sections. Smart objects as the major application domain and one of the driving forces motivating this work are introduced in Section 2.1, which outlines relevant concepts and defines important terms on the background of established terminology of related research areas. With rising importance of the research landscape around ubiquitous computing (UbiComp), human-computer interaction (HCI) is constantly shifting from desktop systems towards everyday environments, which has also coined the notion of human-environment interaction. Thus, longstanding interaction paradigms are being revisited and applied to novel settings that increasingly involve the user's physical reality. We review the current state of interaction design and introduce fundamental concepts that relate to this thesis in Section 2.2. Within the domain of HCI, this work builds primarily on two major disciplines, namely tangible interaction and anthropomorphic user interfaces, which we will introduce and discuss in Sections 2.3 and 2.4.

We conclude this chapter in Section 2.5 with a discussion on the interrelations between these rather detached research fields and examine the potential of combining and integrating specific concepts from these disciplines and its implications for this dissertation.

2.1 Overview, Clarification and Definition of Terms

This section will provide a general overview about the research streams related to UbiComp, a term subsuming a large variety of research activities of a relatively young community. Leaving the era of traditional desktop computing, most terms, concepts and their boundaries in this interdisciplinary research field are still evolving and have yet to be precisely defined and established. We will therefore discuss and order the overall context and particular notions that relate most to our purposes, such that we will be able to define key terms and refine the research context as a background for the remainder of this dissertation.

Particularly the terms smart objects and instrumented environments have been extensively used in a variety of research activities, but a widely accepted definition has not yet been established. In the following we will also review the use of these terms in common literature and identify key attributes of different notions to be able to derive a clear definition of how we use and understand these two terms throughout this work.

2.1.1 Ubiquitous Computing and Related Research Streams

In this section we will introduce and clarify the terms ubiquitous computing, pervasive computing, intelligent environments, ambient intelligence and further related concepts. All these terms are mostly used synonymously to describe the same research area, but a closer look reveals some variations in the foci of their research agendas.

Mark Weiser was the first to use the term **ubiquitous computing** in his often quoted article *The Computer for the 21st Century* where he unfolded his vision about disappearing computers:

> *"The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it."* [Weiser, 1991]

The above statement illustrates the concept of a disappearing technology that is part of our everyday live and activities. Such technology is integrated into the natural environment of humans and blends into the background, with the effect that people do not perceive the digital devices as such and are not aware of their existence when they become integral tools of daily routines. This article was the starting point of the research area UbiComp, which is often synonymously being called **pervasive computing**. The two main conferences are in analogy called *International Conference on Ubiquitous Computing* (since 1998) and *International Conference on Pervasive Computing* (since 2002), attended by basically the same research community. Although the research themes are very similar, one notable difference in the foci of these two research directions is an emphasis on middleware, infrastructures and hardware technologies of the pervasive computing community, while UbiComp takes on a more user-centred perspective, which includes the design of applications, interfaces and their evaluation.

Smart objects belong to UbiComp scenarios and constitute everyday objects enriched with technology - we will examine this term more closely in Section 2.1.3.

The discipline **mobile computing** can also be seen as a subtopic of UbiComp and examines applications on and the technology of personal, portable computing devices such as smart phones or PDAs. As opposed to UbiComp, where users interact with their environment and everyday artefacts, the interaction device in mobile computing is typically a dedicated interface. The boundaries here are of course diffuse, mobile phones are for instance also employed as mediators in pervasive computing scenarios (see e.g. [Rukzio et al., 2006]), in which they act as the user interface to a computationally enriched environment. Research in this area is often concerned about device limitations regarding computational resources and screen space and also covers indoor and outdoor positioning and location-based services. Although mobile computing was originally seen by Mark Weiser as an intermediate step towards UbiComp (later also supported by e.g. Satyanarayanan [Satyanarayanan, 2001]), the recent market penetration of smart phones and the predicted growth amplifies the impact of this research area.**Wearable computing** can be considered as a subarea of mobile and ubiquitous computing and concentrates on technology that is directly worn on the body or embedded in clothes, such that the user can benefit from services and interfaces and at the same time move around and interact freely with the environment.

Recently, the term **cyber-physical systems** emerged, describing the integration of computation and physical processes. Such systems typically consist of networked, multifunctional, autonomous, embedded computers that control physical processes, usually with feedback loops where such physical processes affect computation and vice versa [Lee, 2008]. Although these research efforts conceptually relate to UbiComp, the particular perspective emerged from embedded systems and mainly regards technical issues in this domain. Cyber-physical systems differ from traditional embedded systems mainly because they are designed as a network of interacting elements instead of as single, independent devices, a notion that relates more to topics in robotics and sensor networks.

Intelligent environment is another related term, which was gaining popularity, since the first international conference with the same name was held in 2004. The original notion of this concept was first described in 1998 within the context of the Intelligent Room project at the MIT[1] and defines an instrumented environments as an interactive space with embedded technology [Coen, 1998]. The authors delineate instrumented environments explicitly from UbiComp research, by remarking that instrumented environments do not imply that technology is integrated into everyday objects and that the user has to directly interact with such devices. Instead, the constituted embedded technology remains independent from physical objects, although it is as well highly embedded and distributed, including computers with projectors, cameras, microphones and other sensors. But since this particularity is currently not reflected in the topics covered by the Intelligent Environments conference, the differentiation from UbiComp does not seem to be desirable anymore.

The rather similar paradigm of **ambient intelligence** (AmI) has been envisioned and discussed for the first time internally within Philips and was later introduced in 2001 by Euro-

[1]Massachusetts Institute of Technology, Boston, USA

pean Union funded efforts of the Information Society Technologies Advisory Group (ISTAG) in a report about their vision of information society in 2010 [Ducatel et al., 2001]. In direct comparison to UbiComp and pervasive computing, AmI is less architecture and middleware focused and instead emphasises intelligent functionality, personalised applications and user-friendly interfaces. The key features of AmI are [Aarts and Marzano, 2003]:

Embedded: Networked devices are integrated into the environment

Context-Aware: Devices are able to recognise situational context

Personalised: Devices can be tailored towards the user's needs

Adaptive: Devices can change in response to the user

Anticipatory: Devices can anticipate plans and intentions of the user

This makes AmI closer to UbiComp than to pervasive computing, while emphasising the notion of ambient or invisible system response. This perspective allows us to apply the term *AmI application* as a type of application that covers smart object systems as well as intelligent environment applications.

Finally, the term **instrumented environment** basically denotes a physical environment that is enriched with technology, enabling UbiComp and related scenarios as discussed above. As we will use this term throughout this work, we will discuss and define this term in more detail in the coming section.

2.1.2 A Definition of Instrumented Environments

Pioneering and often cited work in this area was the *Aware Home* Project at the Georgia Institute of Technology. The main intention of this setup was to build up a living laboratory of UbiComp in support of home life, particularly of the elderly [Kidd et al., 1999]. *Context sensing* through an abundance of deployed sensors was the major task in this scenario, trying to sense and understand the activities of the *Aware Home* inhabitants. Thus, the notion of instrumented environment in this project was the sensing capability of the environment to generate a stream of context information enabling new services in a home scenario.

Another interpretation of instrumented environments extends this thought and outlines *invisibility* as another major attribute beside *ubiquity* [Estrin et al., 2002]. Further, the role of *actuation* of embedded devices is introduced as an additional but secondary task besides sensing. The authors also note that the communication between such networked, embedded nodes to support coordinated and organised actions make an important difference between instrumented environments and spaces full of electronic devices.

Stahl et al. state in their work that instrumented environments incorporate distributed computing and sensing devices and also presentation media, which are all used by a variety of ubiquitous services [Stahl et al., 2005]. They emphasise the importance not only of detecting explicit but also implicit user interactions to infer activities as well as intentions in order to build pro-active assistance systems. The authors examined in their scenario an airport environment, a large public space with a potentially large amount of users and output devices that can be either public or private, like wall-sized screens, public audio systems or small, single-user screens [Stahl, 2009].

Schneider considers instrumented environments as parts of the real-world that are equipped with UbiComp infrastructure for the realisation of value-added services [Schneider, 2010]. Such environments are able to interact with the user and/or adapt their physical state through a set of actuators like information displays (including visual, auditory, and tactile displays), or physical actuators. Smart objects are also explicitly mentioned as potential components in instrumented environments, as well as a communication infrastructure for wireless data exchange. It is further stated that the physical elements of the environment are augmented with a virtual layer, linking the physical environment to digital information sources and applications.

We may summarise that the most important key aspect is the ubiquitous and unobtrusive deployment of sensing hardware for detecting user activities with the main objective of providing assistive services in such environments. Looking at the evolution of the term instrumented environments, we can say that some sort of communication, among embedded devices and services, either centralised or distributed, is also a requirement for an instrumented environment. We also like to note that all of the mentioned contributions stress the importance of *localisation* as one of the most important context information for systems in an instrumented environments setup. For our work we therefore define instrumented environments as follows:

Definition 2.1 (Instrumented Environments) *Instrumented environments are spaces of everyday public, professional or private life, enriched with embedded devices, which possess sensing, computing, communication and/or actuation abilities. An instrumented environment is further complemented with output devices such as screens or loudspeakers. The environment hosts services that are able to access these embedded devices to process and evaluate incoming data to derive user activities and other context information.*

2.1.3 A Definition of Smart Objects

The research area of smart objects is a particularly young discipline and is now taking the leap from prototyping and exploring technological dimensions towards concept oriented research. As a subtopic of Pervasive and UbiComp, the aspects of smart objects just recently became an independent research theme on its own, resulting in more focused conferences such as *The Smart Objects and Ambient Intelligence conference* and workshops such as the *Workshop on Design and Integration Principles for Smart Objects* (DIPSO) and the *Workshop on Smart Object Systems*.

Pioneering work in this field by Beigl et al. defined a smart object as *"An everyday* *artefact augmented with computing and communication, enabling it to establish and exchange information about itself with other digital artefacts and/or computer applications.*" [Beigl et al., 2001]. The main point of this definition, which is also reflected in their MediaCup system, is that the object collects and shares information instead of utilising it for independent and autonomous actions.

A part of the Disappearing Computer Initiative (DCI) by the EU was the Smart-Its-Project [Holmquist et al., 2004a], pioneering in developing a platform for embedding sensor nodes into everyday projects by providing a framework of small, networked sensors. From the viewpoint of the DCI, a smart object is an everyday object complemented with digital means of information processing and exchange, based on sensors, actuators and processors. Further, smart objects are capable of communicating with other nearby artefacts and also globally, mainly functioning individually but also as an ensemble.

With the DCI focusing on standalone smart objects, the DIPSO-Community/Organisers take a broader perspective on smart objects and particularly notice the interdependence with instrumented environments - besides the technological requirements of smart objects, which basically aligns with the point of view of the DCI.

Mattern et al. slightly deviate from this viewpoint and emphasise that smart objects are also able to "*...discover where they are, which other objects are in their vicinity, and what has happened to them in the past.*" [Mattern, 2003]. The main difference is the explicit state awareness of an object, which enables e.g. to keep track of interaction histories. The working group around F. Mattern located at ETH Zurich particularly focuses on smart objects in the context of the Internet of Things, basically a network of objects by attaching uniquely identifying devices to objects [Mattern, 2005] [Mattern, 2004] [Bohn et al., 2004]. This term is coined in 1999 by the Auto-ID Labs, a research group consisting of seven institutes (University of Adelaide (Australia), University of Cambridge (United Kingdom), Fudan University (China), Information and Communications University (South Korea), Keio University (Japan), Massachusetts Institute of Technology (USA), University of St. Gallen/ETH Zurich (Switzerland)) in cooperation with EPCGlobal, working on Radio-frequency identification (RFID) technologies and standards towards an architecture for the Internet of Things. The additional notion of smart objects here is the fact that a minimal smart object instrumentation can consist of a RFID-tag for allowing simple identification and in consequence results in immediate economic benefit, typically in logistics or other object tracking scenarios. From this economic point of view an instance of smart objects could also be self-conscious product that perceives and observes its condition, state changes, attempting to change the situation if not satisfied with it, which would in turn reduce total costs of making procuring replacement more efficient or even avoiding it completely. Such technology enables the realisation of various applications, which are for instance being developed in the SemProM (Semantic Product Memory) project [Wahlster et al., 2008], which focuses on utilising product histories in order to facilitate value-added services in different stages of a product's life-cycle.

Another definition of smart objects was proposed by Kortuem et al. in 2007, outlining them as "*Objects of our everyday lives, augmented with information technology and equipped with sensing, computation and communication capabilities, that are* *able to perceive and interact with their environment and with other smart objects.*" [Kortuem et al., 2007]. The emphasis here lies on the ability of collecting and communicating context information, while the properties of the original object such as appearance and use are retained.

For this work we prefer to use a more comprehensive notion of smart object, such that this term does not necessarily imply intelligence or processing on the object itself, but includes the possibility of that. We even consider objects as "smart", when technological enhancements are not physically embedded into the actual device but distributed in the environment, in a way that an unambiguously identifiable object still obtains unique additional qualities as discussed above. This perspective also includes artefacts that are merely augmented with identification and possible tracking technologies such as visual markers or RFID tags. A borderline case to smart objects are certain types of robots, which are designed to either appear similar to or to fulfil functions of a physical object that already exists. For example a robotic vacuum cleaner that is able to autonomously clean ground floors, avoiding obstacles through its sensory perception. While typical robotic systems are not modelled after existing devices but designed for a dedicated self-contained purpose, we may consider this class of robots as smart objects as well as long as the original function is preserved, e.g. the manual use of the aforementioned robotic vacuum cleaner.

Based on these considerations and previously mentioned definition proposals, we define smart objects for this work as follows:

Definition 2.2 (Smart Objects) *Everyday objects augmented with identification technologies and/or sensing and computation power, that potentially communicate with the environment, other objects and/or the user. The original appearance and functionality of the object are retained as much as possible. Such objects enriched with a new quality are either self-aware and independent actors or interact with an instrumented environment to facilitate user assistance.*

2.2 Interaction Design

The demands of interaction design have substantially changed during the past decade with the advent of application scenarios related to UbiComp. The previously dominating area of graphical user interface (GUI) design is now only one of several topics on the increasingly diverse agenda of HCI research issues. In this section we will briefly introduce interaction design in general and elaborate relevant concepts that have emerged with the inclusion of physical realities and ubiquitous technology into the design space.

2.2.1 Interaction Design and Qualities of Interaction

Interaction design in general is concerned with the design of function, behaviour and appearance of products and systems, which includes all aspects that are perceived by the user while interacting with the application or device. While interface design is historically associated with the design of graphical user interfaces, the discipline of interaction design particularly includes the definition and presentation of interaction behaviour over time. This very interdisciplinary research area requires expertise in several disciplines, including, for instance, sound design and industrial design (see Figure 2.1).

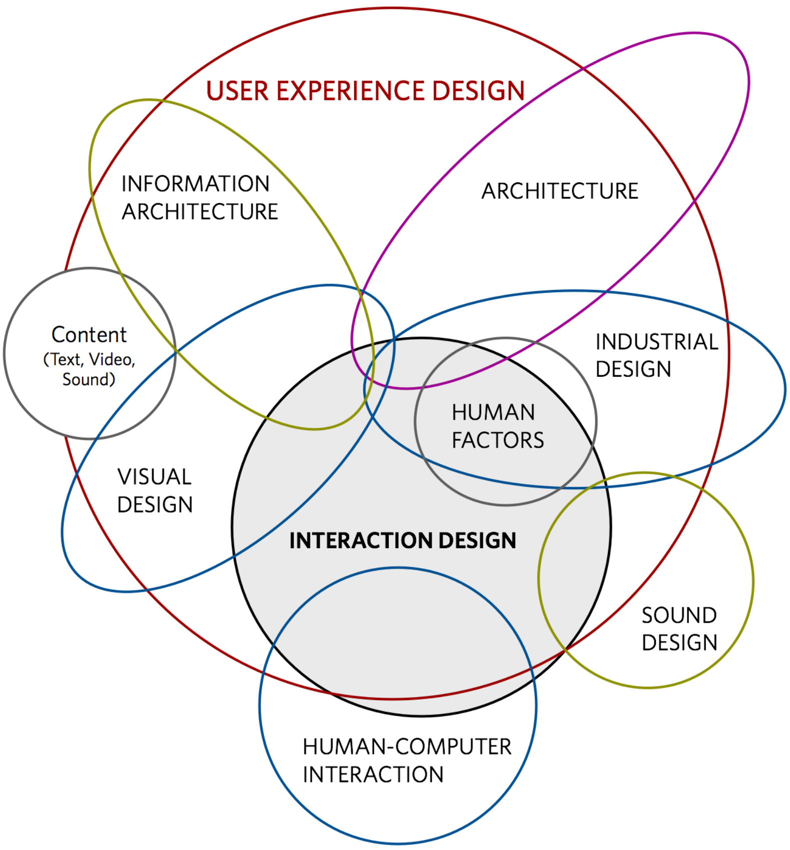


Figure 2.1*: Disciplines involved in interaction design (source: [Saffer, 2009])*

There are several attempts to define interaction design from different perspectives: For instance Marsden et al. have discussed this term in the context of mobile computing and concluded with the following [Jones and Marsden, 2006]:

> *"Interaction design creates a plan specifying the user needs in terms of required functionality, how this functionality is to be accessed and controlled, the presentation of content, system state, help and feedback information, and the way the system is to integrate with other resources in the user's context."*

This rather technical perspective reflects the traditional point of view on interaction design that focuses on efficiency and *usability*. Usability, also referred to as *ease of use*, is basically concerned with task efficiency the absence of usability faults and avoidance of frustration of the user while using the system. Various definitions with different notions exist. Winograd for instance defined interaction design in a broader sense and created an analogy to architectural design of spaces for human communication and interaction: From his point of view software is creating spaces, in which users live and act - just as in constructed real spaces that provide possibilities and options to the inhabitants that occupy this space and fill it with life [Winograd, 1997]. Whereas Shedroff described interaction design as an art to design valuable, meaningful, interesting, convincing information, experience and interaction with user control and feedback as essential components [Shedroff, 1999]. This notion addresses additional facets beyond usability, including emotional qualities, also described as *joy of use*, *hedonic quality* or *interaction experience* - we will stick to the term joy of use throughout this dissertation. Accordingly, discussions in recent years have raised criticism towards traditional human computer interaction research for primarily addressing cognitive aspects of the user and neglecting the physical and emotional ones. User experience and joy received more attention and became criteria for the evaluation of interactive products and increasingly a topic in interaction design research [Hassenzahl et al., 2001][Monk et al., 2002]. This paradigm shift is well expressed by a statement of Djajadiningrat [Djajadiningrat et al., 2000]:

> *"A user may choose to work with a product despite it being difficult to use, because it is challenging, seductive, playful, surprising, memorable or rewarding, resulting in enjoyment of the experience."*

While ease of use aims at avoiding user frustration during the interaction, joy of use motivates to (re-)use the product or system. In order to setup criteria for evaluating interfaces, such as described in more detail in Section 2.2.2, traditional usability measures certainly have to be addressed to evaluate ergonomic goals of the system but also qualities of an interface that go beyond usability, which have previously been disregarded. As this research area has just begun to explore the design space, it is not possible to determine a definite list of criteria, instead we present and analyse the current state of art and align our efforts accordingly.

2.2.1.1 Ease and Joy of Use

The discipline usability (or ease of use), delivers norms and standards about, as well as theories and concepts of design, analysis and evaluation of software interfaces. The main goal is to create user-friendly human computer interaction by designing the system depending on the user's abilities. Usability has been defined in 1998 as part of the international norm ISO 9241-11 (Ergonomics of Human System Interaction) as follows [ISO/IEC, 1998]:

> *"The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use."*

The key terms of this definition are specified as follows:

- **Context of use**: Describes the user and the environment in which the product is being used
- **Effectiveness**: Describes the accuracy and completeness of the goal achievement of the user
- **Efficiency**: Describes the relation between the accuracy / completeness and the required efforts of the user to achieve the specified goals
- **Satisfaction**: Describes the degree of freedom from restrictions and includes the acceptance of and positive attitude towards the product

It has been discussed whether the latter factor (satisfaction) does already cover joy of use, but the ISO definition only states that a user is positive towards a product that is efficiently and effectively designed. Satisfaction due to enjoyment in a hedonic sense is not mentioned. Therefore the phenomenon of joy of use is at least partially not covered by the usability paradigm according to ISO 9241-11 [Djajadiningrat et al., 2002].

In accordance to this definition the related norm ISO 9241-10 is concerned with general ergonomic design principles of dialogues between human and information systems, which we will briefly specify in the following paragraph [ISO/IEC, 1996]:

- **Suitability for the task**: A dialogue is suitable for the task, if it provides assistance in order to fulfil the tasks effectively and efficiently.
- **Self descriptiveness**: A dialogue is self descriptive, if each step of the dialogue is immediately comprehensible due to system feedback and without requiring assistance of a third party such as a user manual.
- **Controllability**: A dialogue is controllable, if the user can control the start of the dialogue and also the direction and speed until the goal is reached.
- **Conformity with user expectations**: A dialogue is conform with user expectations, if it is consistent and complies with the characteristics of the user.
- **Error tolerance**: A dialogue is error tolerant, if the intended request can be accomplished with minor correcting efforts after erroneous inputs.
- **Suitability for individualisation**: A dialogue is suitable for individualisation, if the system allows to make adjustments to the requirements of the task as well to the individual preferences of the users.
- **Suitability for learning**: A dialogue is suitable for learning, if support is provided to learn the use of the dialogue system.

Thus, the ease of use and resulting design principles have been internationally standardised and recognised, while the concept joy of use, its definition and possibilities of evaluation are still unclear. A common understanding of which specific requirements have to be fulfilled to achieve this effect at the user is not yet given. But several research groups have examined and expressed criteria that are crucial to create a joy of use (see e.g. [Overbeeke et al., 2004] or [Brandtzaeg et al., 2004]. Since this research discipline is exploring new grounds, a common denominator has yet to be established from scientific discourse. Reeps has analysed and summarised previous efforts of determining quality criteria of joy of use and attempted to unify parallel research results in her work [Reeps, 2004]. The relevant key factors can be outlined as follows [Carsten, 2009]:

- Appropriate **challenge**
- Provoking **curiosity**
- Allowing **personal experiences** of use
- Allowing for **social interaction** and interaction as motivational factor
- Pleasant **visual design**
- Simple, **intuitive and inspiring** application
- **Control**
- **Individualisation**
- Clear **feedback**

It has further been acknowledged that usability is a base requirement for joy of use, because a lack of usability would result in dissatisfaction, frustration or other negative emotions, before a joy of use can develop. An enjoying user experience would be hardly possible. The latter three issues in the list of key factors are already part of the usability paradigm, underlining the importance of their presence for the factor joy of use and the blurring boundaries between these two concepts.

Hornecker has derived further quality criteria and desirable attributes particularly for tangible user interfaces (TUIs) (see also Section 2.3) from contemporary literature and discussions, constituting additional valuable directions for this work [Hornecker, 2004]:

- **Affordances** (see Section 2.3.2 for a definition) should be clear and easily recognisable
- Interacting with the system should be as **natural and intuitive** as interacting with the real world
- **Complexity** of the real world should be reduced not reproduced
- **Directness and enjoyment** of use

- **Physical motion** should be supported, since it perceived as activating and invigorating.
- Systems should sometimes be **surprising** or offer **resistance** instead of completely complying with expectations, also called the quality of surprise in [Löwgren, 2001]. This provokes curiosity and can support exploration and learning - although this could potentially contradicts with the ergonomic design principle of conformity with user expectations [Crutzen, 2000].
- Appropriate **balance of material and virtual representation**, such that on the one hand real artefacts do not appear unimportant or even superfluous or disturbing and on the other hand virtual components do not appear to be not contributing to the function or not being registered by the user's awareness.

The latter issue was already formulated by Ullmer and Ishii in their summary of the current state of TUI research in 2001 [Ullmer and Ishii, 2001]:

> *"Such systems require a careful balance between physical and graphical expression to avoid physical clutter, and to take advantage of the contrasting strengths of different representational forms. This balance between physical and digital representations stands as one of TUI's greatest design challenges."*

Both sets of criteria, the key factors of joy of use and desirable attributes for TUI are set to guide the development of user interfaces in this work.

2.2.2 Post-WIMP Interaction

> *"From the isolation of our workstations we try to interact with our surrounding environment, but the two worlds have little in common. How can we escape from the computer screen and bring these worlds together"* [Wellner et al., 1993b]

Research areas in the periphery of UbiComp as introduced in Section 2.2.1 exhibit additional dimensions for the design space of interfaces that are known from previous, GUI-based desktop interaction paradigms. The so called WIMP[2] metaphor has been the most prominent interaction model and can be outlined as follows [Beaudouin-Lafon, 2000]:

- Application objects are displayed in document windows
- Objects can be selected and sometimes dragged and dropped between different windows; and

[2]WIMP = **W**indows, **I**cons, **M**enus and **P**ointer

- Commands are invoked through menus or toolbars, often bringing up a dialog box that must be filled in before the command's effect on the object is visible.

Although such a paradigm is theoretically transferable, WIMP does not address the needs and peculiarities of UbiComp applications, therefore the question of ergonomic interface design in UbiComp or tangible interaction has been increasingly raised in recent years. Experiences of GUI-based design can in most cases not be applied to new interface requirements.

Post-WIMP interfaces have been defined by van Dam as interfaces "*containing at least one interaction technique not dependent on classical 2D widgets such as menus and icons*" [van Dam, 1997]. This includes virtual reality, mobile computing and affective computing as well as interfaces in UbiComp scenarios.

Most prototypes so far were developed in an opportunistic manner, as noted by Dourish [Dourish, 2001], p. 52, initiated and inspired through the development of new sensor technologies or input devices. Concepts and aesthetics being secondary, the primary goal was to explore the design space and to discover the potential of new technologies.

The need for alternative solution for GUIs accompanied by conceptual ideas was published for the first time in a 1993 in the CACM journal issue titled *"Back to the Real World"* [Wellner et al., 1993a]. It has been argued that humans live and work in the real world that surrounds them, with their abilities to interact and move within that well developed. But the work at a computer reduces their actions to hand-eye coordination tasks with screen, keyboard and mouse. The computer therefore isolates the user from the social environment and the real world, which makes computer workstations and the remaining environment two disjunct worlds that have to be synchronised with additional efforts from the user. Switching between computer based to traditional media requires a change of tools, environments and interaction styles, the main intention of new interaction paradigms is therefore to merge these two worlds. As also Weiser stated in [Weiser, 1991], the aim is to let the user remain in her familiar environment, where she is able to experience with all his senses, to communicate with others and to maintain known interaction styles for objects and information. This is particularly opposed to the direction of *virtual reality*, which separates users from reality, in order to exchange it with a computer generated world. But the real, material world should not be replaced but augmented, everyday objects should be extended with digital features but keep their familiar attributes. The natural interaction with objects and other humans should not be encumbered, interaction with augmented artefacts takes place in context, supporting situated interfaces. Another important trend in this context is denoted by the term *dual reality*, which is defined by Lifton as an environment that results from the interplay between the real world and the virtual world [Lifton, 2007]. While both worlds are complete by themselves, they have the ability to "*mutually reflect, influence, and merge into each other by means of sensor/actuator networks deeply embedded in everyday environments.*"

Conceptual work in this area has only recently emerged, the fundamental steps to unify and classify novel, post-WIMP interaction paradigms will be introduced in the following sections.

2.2.2.1 Embodied User Interfaces

> *"So, why can't users manipulate devices in a variety of ways - squeeze, shake, flick, tilt - as an integral part of using them?"*[Fishkin et al., 2000b]

The idea of *Embodied User Interfaces* suggests to accept that computing power is embedded into physical devices, such that the manual interaction with the body of the device should be an integral part of the interaction [Fishkin et al., 2000b]. Fishkin criticised at that time that in mobile computing interacting with small, ubiquitous devices still follows the same paradigm established for desktop computers, only that the mouse as a pointing device was replaced by a stylus. He proposes an interaction style that mimics non-digital actions, which are known from the analogue version of a device and that the user is familiar with. He provides examples and prototypes, for instance a tablet PC for document viewing that flips through pages if the user flicks on the upper corners of a the device - an action that is similar to flipping papers of a real book. The consequence of this notion would be a paradigm for UbiComp interfaces that conform with the ideas of tangible interaction (see Section 2.3). The key point here is to regard the physical, analogue use of a smart object in the design of extended, digital functionality. Fishkin continued his research on the idea of embodiment and formulated a taxonomy several years later (see Section 2.3.1).

2.2.2.2 Reality-Based Interaction

The concept of *Reality-Based Interaction* (RBI) has been proposed and published at CHI'08[3] as a unifying concept for the wide variety of emerging post-WIMP interaction styles [Jacob et al., 2008]. This framework is supposed to be used to understand and compare current paths of HCI research, also in order to identify gaps or opportunities for future research, and to analyse and relate specific interaction designs. The argumentation starts with the observation that such interfaces commonly build on users' pre-existing knowledge of the everyday, non-digital world to a much greater extent than previous interface paradigms such as WIMP. Jacob et al. argue that there are four core themes of the physical, non-digital world, which are exploited by these novel interaction designs, because they enable users to capitalise from their real world skills in human computer interaction, reducing the required mental effort. The authors claim that this may speed up learning, improve performance and foster improvisation and exploration, since users do not need to learn proprietary, interface-specific skills. These four RBI themes are described as follows [Jacob et al., 2008]:

- Naive Physics: People share a common sense knowledge about the physical world, including concepts such as gravity or velocity, which are often simulated or exploited. A TUI might, for example, apply physical constraints to suggest the way in which a physical object can be manipulated (as in Section 2.3.2).
- Body Awareness & Skills: people have awareness and control of their own physical bodies, independent of the environment. For example, a person is aware of the relative

[3] The Twenty-Sixth Annual SIGCHI Conference on Human factors in Computing Systems

position of his or her limbs. Early in life, most people develop skills to coordinate movements of body parts such as limbs, head, eyes, and so on, in order to crawl, walk, or kick a ball. Several emerging interfaces support increasingly rich sets of input techniques based on these skills, including two-handed interaction and whole-body interaction.

- Environment Awareness & Skills: people have a sense of their surroundings and possess skills for negotiating, manipulating, and navigating within their environment. Clues, e.g. landmarks, that are embedded in the environment facilitate our sense of orientation and spatial understanding

- Social Awareness & Skills: People are generally aware of the presence of others and develop skills for interacting with them (socially). These include verbal and non-verbal communication, collaboration or the ability to exchange physical objects. For example, TUIs provide both the environment and input devices to support co-located collaboration. Virtual environments build on social awareness and skills by representing users' presence with avatars.

Notably, the concept and potentials of anthropomorphism have not been discovered and regarded in this work, although it can be theoretically subsumed by the theme ”Social Awareness & Skills”, which is one important requirement of certain anthropomorphic interfaces. A reason might be that the notion of life-likeliness has not yet been consequently applied to interfaces in physical, reality-based interaction, underlining the novelty of the work at hand.

The authors generally propose that interfaces should attempt to be as close as possible to reality in the chosen themes. But it is also outlined that tradeoffs exist and that in some cases the degree of reality might better be reduced to trade for other important qualities, noted as:

- Expressive Power: The variety of tasks that can be accomplished within the application domain

- Efficiency: The possibility to perform tasks rapidly and reach goals in a short amount of time

- Versatility: Enabling users to solve tasks from different application domains

- Ergonomics: Allowing users to complete a task without physical exertion or fatigue

- Accessibility: Allowing a wide range of users with a variety of abilities to complete a task

- Practicality: Ensuring the development and production of the system at reasonable costs

Some of these qualities are either equivalent to aspects of usability as defined in ISO 9241-10 or at least partially resemble their intention: E.g. efficiency here is basically equivalent to *suitability for the task*. Ergonomics in this context regards physical concerns and is also subsumed by suitability. Accessibility and expressive power partially overlap with suitability, but go further and explicitly mention new concerns. Whereas practicality introduces the issue of realisation costs, a concerns that is certainly important but normally not considered in a research context or design oriented point of view. Noting these similarities to usability, it is not elaborated how hedonic qualities as in joy of use (see Section 2.2.1.1) should be integrated into the RBI concept.

The stated general guideline for the design of an RBI interface could be summarised as: Develop an application according to the RBI principles, avoiding the mentioned tradeoffs if possible. Otherwise analyse, discuss and integrate tradeoffs as described in the paper.

2.3 Tangible Interaction

The trend of recent years towards shifting computing into the real world evolved and manifested in several research streams such as UbiComp and tangible interaction. As elaborated in Section 2.1.1, the overall common goal was to augment and enrich the physical and familiar world with new services instead of replacing it as in virtual reality efforts. Established human skills and knowledge should be exploited as a ground for new interfaces.

The fundamental principle of a TUI is - as the name implies - the tangibility of the interface and the material form of a digital resource [Hornecker and Buur, 2006]. In this manner the interaction with our everyday environment is maintained and the interface is situated into real world context. This has the effect that users are enabled to utilise their experiences with the real world, the threshold for activity is lowered and the access bottleneck of the keyboard bypassed [Stanton et al., 2001]. This notion refines the intuitive understanding of a TUI, which previously describes TUIs simply as interfaces that can be touched or grasped - which would also hold true for keyboard and mouse. Before we introduce relevant concepts, we will have a brief look at the comparably young history of this research stream to be able to understand of the current state of art in this HCI domain.

Some first TUI prototypes appeared in the domain of architectural planning in the early eighties with the goal to improve existing CAD-Applications, mainly in order to be able to involve clients in the design phase of buildings and to assist architects and clients in their negotiations about design alternatives [Aish and Noakes, 1984] [Frazer et al., 1980] [Frazer et al., 1982]. Due to limited technologic possibilities at that time, these prototypes were rather bulky and inconvenient, which might have been a reason why these TUI predecessors remained quite unpopular. Later, around 1993 the general idea revived within the HCI community and was published by different groups in different notions: The widely known Marble Answering Machine [Poynor, 1995] [Abrams, 1999] sketched by the industrial designer Durell Bishop in 1992 is often quoted as an inspiring example. The Marble Answering Machine physically instantiates incoming messages as marbles, which can be picked up and placed into a small indentation of the answering machine to listen to that message. To call back the person who left the message, the user can place the marble in the call-back indentation of the phone. In 1995 Fitzmaurice, Ishii and Buxton created the “”Graspable

Bricks“” [Fitzmaurice et al., 1995] and thereby coined the term *Graspable User Interfaces*. Ishii and Ullmer published the most wide spread definition of *Tangible User Interfaces* in 1997 during their research efforts within the Tangible Bits project [Ishii and Ullmer, 1997], which started in 1995. The basic idea shared by these variations was to use physical objects to represent or control digital information. The essential characteristic of a TUI according to [Ishii and Ullmer, 1997] is the seamless integration of representation and control, with physical objects being both representation of information and as physical controls for directly manipulating their underlying associations, both input and output device at the same time. There are four key characteristics regarding representation and control:

1. Physical representations are computationally coupled to underlying digital information.

2. Physical representations embody mechanisms for interactive control.

3. Physical representations are perceptually coupled to actively mediated digital representations. (e.g. visual augmentation via projection, through sound...)

4. Physical state of tangibles embodies key aspects of the digital state of a system.

This implies that TUIs rely on a balance between physical and digital representations: While physical representation (e.g. spatial configuration) remains meaningful in itself, digital information is needed to mediate dynamic information. In Fitzmaurice’s approach [Fitzmaurice, 1996] on Graspable UIs, a focus lies on spatial arrangement and direct manipulation. From his point of view Graspable UIs involve multiple input devices serving as dedicated physical interface widget affording physical manipulation and spatial arrangement, leveraging off of a user’s everyday knowledge such as spatial reasoning skill and physical object manipulations. The possibility to associate devices independently with different functions allows to take advantage of shape, size and position of physical controllers to increase functionality and decrease the complexity of use, making such devices to ”graspable functions”.

The common goal is simple: Creating an intuitive user experience by turning our physical world into an interface, by bringing digital information to the familiar real world, where it turns visible and manipulable. These concepts and the variety of prototypes have shown that there are two key attributes of TUI systems that are originally inspired by Bishop’s Marble Answering Machine and repeatedly discussed [Abrams, 1999] : **Affordance** and **Mapping**. Affordance of a virtual or physical object according to Norman is a quality of the object (e.g. its shape) suggesting the way of using it, a concept that is often utilised by designing shape to enable intuitive use. Good affordances therefore yield an effective elevation of human everyday knowledge and skills regarding physical manipulation (Affordances are discussed in more detail in Section 2.3.2). Nevertheless, critics argue that the concept of affordances is too limited, since it does not consider emotions and expectations of users, thus the aesthetic and emotional components of interaction are disregarded (see also Section 2.2.1).

Mapping in the context of human-computer interaction generally constitutes the perceived relationship between interface elements and their movement on the one hand and their impact on the target system on the other hand. *Spatial mapping* (as in e.g. [Sharlin et al., 2004]) particularly regards the location and spatial arrangement of interfaces and their meaning in respect to their function and type of interaction. A natural mapping (according to [Norman, 1998]) for instance utilises spatial analogies and cultural standards. Another related concept that also emerges in the context of the Internet of Things is *object mapping* - the association of new meaning to arbitrary objects. Dedicated interface objects can be mapped to functions and controls whereas everyday objects can be extended with digital properties and point to any kind of digital information.

TUI research merely focused on developing new prototypes until recently, when a shift towards theory and frameworks can be noticed such as in the special issue on *"Tangible Interfaces in Perspective"* of the journal on *"Personal & Ubiquitous Computing"* [Holmquist et al., 2004b]. These attempts so far resulted in taxonomies defining and categorising systems (e.g.) [Fishkin, 2004] [Sharlin et al., 2004] [Benford et al., 2005] [Hurtienne and Israel, 2007] and outline the design space or unexplored territory.

Recent attempts aimed at generalising previous categorisations from a pure TUI paradigm towards embodied, situated interaction for UbiComp scenarios, most notably Reality Based Interaction (see also Section 2.2.2.2).

2.3.1 Embodiment and Metaphor - A TUI Taxonomy

An essential framework in the TUI realm was formulated by K. P. Fishkin, integrating and unifying previous categorisations and definitions [Fishkin, 2004]. Another reason for us to consider particularly this taxonomy over other frameworks is its object centric perspective matches with our approach on smart object interaction - as opposed to, for instance, integrating movement and space as in [Hornecker and Buur, 2006].

Fishkin identified the two concepts *embodiment* and *metaphor* as the axes in the design space of TUIs and argues that a high level in each attribute increases the *tangibility* of the interface, reducing learning efforts, increasing the impact of the metaphor and thus adding value to the whole interactive system.

Embodiment specifies how closely tied the input focus to the output focus is. A high amount of *embodiment* would therefore imply that the user perceives the state of computation as *inside* of the manipulated device, embodied within a physical housing. This characteristic is further classified into four discrete levels, depending on how the output of an action spatially coincides with the input device: *full*, *nearby*, *environmental* and *distant*.

Full: In this case, the output device equals the input device, such that the state of the system is fully embodied in the device. An analogy is clay sculpting, where clay is pushed and at the same time the effects can be seen on that particular clay. This level of embodiment is the most common one from the physical world.

Nearby: For this level of embodiment, the output takes place near the input object. Although not within the same object, the output is still tightly coupled to the focus of the input. For example, a light pen that modifies the display directly beneath it, would express this level of embodiment.

Environmental: In this case, the output is somewhere in the environment of the user, for example audio rendered by loudspeakers in the room, or changes of light or heat levels. Although there is a logical link between the input object and the source of output, the output is perceived as separated from it.

Distant: For distant embodiment, the output appears separated from the object, e.g. on another, distant screen or even in another room. For example a TV remote control, which triggers a switch of the visual attention between input device (remote control) and the output (the TV screen).

The *metaphor* dimension describes the degree of analogy of a system affect to a real-world effect of similar action. This dimension is further grouped into two types, the *metaphor of noun*, those which appeal to the shape/look/sound of objects, and *metaphor of verb*, those which appeal to the motion of objects, i.e. the analogy of the interaction to real world actions. The more of one type of metaphor is used the higher the scale on the metaphor rating.

In particular, if *no metaphor* at all is employed, users typically apply physical manipulations for controlling a system, but these manipulations contain no analogies to any real-world actions. For example, a command-line interface where typing gestures have no correlation to the effect, e.g. copying a file on the local hard-drive.

The metaphor of *noun* relates to systems, which employ an analogy to the outer appearance of an object in the real world. But, the actions performed on/with that object are not analogous. An example from conventional HCI would be windows desktop systems, where file or directory icons on virtual desktops were analogous to physical pieces of paper or drawers on physical desktops. However, most physical actions on real desks (clean, crumple, staple) have no virtual counterpart, and many of the virtual operations have no physical analogy (e.g. resize). In the field of TUI, this level is reached by systems in which merely the look of an input object resembles the look of a real-world object.

Accordingly, systems might apply an analogy to the actual gesture or movement performed in a real world scenario, which is called an analogy of *verb*. The visual appearance of involved objects is irrelevant and is connected to a real world counterpart.

In an analogy of *noun and verb*, both appearance and action are now related, but the physical and virtual objects still differ. In our windows desktop example, drag-and-drop applies this level of embodiment, since dropping an icon into a virtual wastebasket refers to dropping physical paper into a physical wastebasket.

A *full* metaphor means that the user does not need to make an analogy at all, the virtual system appears to be the physical system, which implies that manipulating the object changes the world in the desired way. Although a high rating in the proposed design spectrum is generally desirable in the spirit of tangible interaction, design trade-offs exist and have to be considered.

Taking the perspective of this taxonomy also outlines trends in the evolution of TUI prototypes. An examination of over 60 TUI systems has shown that there has been a clear trend of increasing the levels of embodiment and metaphor in the development life-cycles [Fishkin, 2004].

Implementing interfaces for smart objects can employ both a high degree of embodiment and metaphor of noun, due to their inherent physicality combined with embedded technology. Furthermore, tangible input and anthropomorphism have high potentials for realising a high metaphor of verb. And in consequence enabling the achievement of both full metaphor and full embodiment.

2.3.2 Affordances

The term affordance is originally introduced by the psychologist James J. Gibson [Gibson, 1977] and defined as the quality of an object or an environment, that allows an individual to perform an action. Donald Norman introduced the term into the HCI area and refined its meaning for this domain [Norman, 1999]:

> *Affordance of a virtual or physical object is a quality of the object (e.g. its shape) suggesting the way of using it, which is therefore an intuitively recognised relation between attributes of an object and possible actions or operations, depending on the physical capabilities, goals, plans, values, beliefs and past experience of the actor.*

Affordances are calls or options for action and describe properties of environments or objects that are relevant and meaningful for persons, because they allow certain activities or operations. They therefore describe a relation between those attributes and possible interactions and indicate how to perform certain actions. For instance, a door handle invites to push it down or a knob to be turned [Norman, 1998]. Norman combines the concepts Mapping, Visibility and Affordances to a design theory. Mapping constitutes the perceivable mapping between intended action and result or process and between interface elements and effect on the system. So called natural mapping utilises analogies, e.g. spatial, and cultural standards. Visibility means that components of a system are visible and convey an appropriate message. But affordances are not always visible, known or wanted, they can for instance also be discovered through exploration and experimenting with the device.

Criticism shows that the concept of affordances is not yet finalised: Its current use by the HCI community is restricted to perceptual and motoric abilities of users, while emotions and intentions are basically excluded. Norman later further differentiated cultural affordances, which are based on common cultural knowledge and learned conventions that can be achieved through active participation. He also noted the difference of *perceived* affordances, which exist particularly in screen-based systems [Norman, 1999]: A desktop computer with display, keyboard and mouse actually affords pointing, touching, looking and clicking on every pixel of the screen and although a nearby screen affords touching, most do not detect and respond to it. Icons, cursors and other graphical elements on the screen only promote *perceived* affordances, using cultural and social conventions and giving feedback. Perceived affordances do therefore also depend on context, culture and the mental model of the actor, while real affordances are independent of the actor's ability to perceive it and do not change as the needs and goals of the actor change [McGrenere and Ho, 2000]. Real and perceived affordances as well as feedback are independent design concepts and can therefore be manipulated independent of one another. Graphical user interface designers do therefore mainly care about perceived affordances. According to Norman, actions of the WIMP paradigm are too abstract and arbitrary in comparison with real, physical manipulation of objects, where the power of real and perceived affordances lies.

2.3.3 Organic User Interfaces

Organic User Interfaces (OUI) as a recent trend constitute a special case of TUIs, which has been defined and discussed in a special issue of the CACM journal in 2008. Holman and Vertegaal [Holman and Vertegaal, 2008] have approached the term by arguing that the planar rigidity of LCD screens is a fundamental limitation for the shape of computing devices and therefore also a limit for their usability and affordances. Whereas paper as a real world displaying device is mentioned as a counter example that can be folded or rolled. They further pointed out that the above mentioned restriction is going to disappear with the emergence of OLED technology and electronic ink, which will allow for the design and construction of flexible and organic-shaped computing devices. Another important development as a prerequisite for organic interface design is that of new materials such as memory alloys, which can be used to mimic the behaviour of muscles and therefore allow for the implementation of shape-changing artefacts with minimised actuators.

Based on these observations, Holman and Vertegaal provided the following definition of Organic User Interfaces [Holman and Vertegaal, 2008]:

> *"An Organic User Interface is a computer interface that uses a nonplanar display as a primary means of output, as well as input. When flexible, OUIs have the ability to become the data on display through deformation, either via manipulation or actuation. Their fluid physics-based graphics are shaped through multitouch and bimanual gestures."*

Three design principles for OUIs are developed in this work [Holman and Vertegaal, 2008]:

1. **Input equals output:** While in traditional GUIs input and output are spatially separated (mouse/keyboard and screen), input should not be distinguishable from output in

OUIs: Users should literally manipulate and deform the virtual objects on the display itself. Such displays should sense their own shape as input, as well as all other forces acting upon them. This principle basically affirms a high degree of embodiment as pointed out in Section 2.3.1.

2. **Function equals form:** The shape of an object should allow a person to perceive what can be done with it. OUIs should physically represent the supported activities, which coincides with the idea of *affordances* as in Section 2.3.2.

3. **Form follows flow:** An attribute that is rather unique to OUIs, due to their shape-changing abilities: An OUI should adapt its form to follow the flow of user activities in a variety of physical and social contexts of use. A form should always regard the activity and change appropriately.

Examples of OUIs are discussed in Section 3.3.3, most of which resemble examples that mimic a notion of living or animalistic artefacts.

2.3.4 Multimodal User Interfaces

Wasinger summarised several definitions and notions of the term *multimodal user interfaces* in his dissertation [Wasinger, 2006]. He outlined that modalities are being referred to as to the *”the human senses which allow incoming information to be received and processed”* [Wahlster, 2006]. In consequence, multimodal interaction is defined to be *”the means for a user to interact with an application using more than one mode of interaction, for instance offering the user the choice of speaking or typing, or in some cases allowing the user to provide a composite input involving multiple modes”* [Froumentin, 2004]. In this sense Schomaker et al. differentiate between interaction that uses only one modality (unimodal interaction), exactly two modalities (bimodal interaction), and two or more modalities (multimodal interaction) [Schomaker et al., 1995]. Multimodal interfaces as a category therefore subsume TUIs, but as a research discipline it focuses on issues that relate to the combination of modalities, such as fusion and fission. Multimodal fusion, or also referred to as multimodal integration is defined in [Froumentin, 2004] as the process of combining inputs from different modalities to create an interpretation of composite input. The main goal of modality fusion is to combine multiple modality input streams into a single result that is modality independent but enriched with semantic meaning. The counterpart modality fission is responsible for splitting such semantic, modality-free meaning into different modality streams for presentation to a user.

The haptic channel on the input side is often used within the process of modality fusion for the disambiguation of references in dialogues, which is also called *reference resolution*. For example, the utterance ”What is the price of this camera?”, occurring in the speech modality, contains the referring expression ”this camera”. Combined with a pointing gesture that refers to a certain product, the referring expression can be matched with this object and the fusion component would be able to infer that the user has inquired the price of that particular product.

2.4 Anthropomorphism

Anthropomorphism is the tendency of people to think of objects as if they had human characteristics. A common and universal definition of the term Anthropomorphism can be found in Webster's New Collegiate Dictionary:

"The Attribution of human characteristics to nonhuman things or events."

One of the earliest documented and discussed instances of this phenomenon occurred when people modelled gods after themselves. This tendency to perceive the world as similar to ourselves was classified as a clear mistake by many scientists, philosophers and others, considering it to be a behaviour which has to be eliminated wherever possible. It is argued that in our epistemological relation to reality Anthropomorphism should be minimised due to its inherent potential faultiness, but anthropologists like Guthrie recognise that it cannot be eliminated completely for being a result of an involuntary and at the same time necessary perceptual strategy of human nature [Guthrie, 1997] [Guthrie and MyiLibrary, 1993]. In this sense it has further been argued that the anthropomorphic mental model may result from a cognitive default [Caporael and Heyes, 1997], resulting from the human mind's powers of discovering and applying analogies and patterns, in this case between ourselves and other phenomena. Thus for the first time a positive connotation of anthropomorphism came into play. This cognitive automatism was further analysed and questioned, scrutinising possible reasons behind this behaviour. Two widespread views attempted to explain the discussed phenomena [Guthrie, 1997]: "According to the familiarity thesis, we use ourselves as models of the world, because we have good knowledge of ourselves but not of the nonhuman world and, looking for an explanation of the world, resort to the knowledge that is easiest and most reliable." In comparison to this cognitive derivation, the comfort thesis refers to the "emotional motive that we are mistrustful to what is nonhuman but reassured by what is human". But Guthrie himself argues that anthropomorphism is a cognitive strategy in the face of chronic uncertainty about the nature of the world. According to him, guessing that some thing or event is humanlike or has a human cause constitutes a good bet, because if we are right, we gain much by the correct identification, while if we are wrong, we usually lose little. Although called a "bet" this is not a deliberate or explicit choice, instead it happens unconsciously as most of perceptual processes do, a product of natural selection, not of reason. It is a strategy to make sense of our world when we are not able to fully comprehend its complexity.

When considering other humans and animals, we reflect on hidden causes of observed behaviour, make attributions as to the traits, experiences, or reasons determining it, and extrapolate to new situations [Povinelli and Bering, 2002]. These tendencies are well established neurologically, and are likely triggered automatically as well when we observe machines that have human attributes and move and speak purposefully [Scholl and Tremoulet, 2000].

2.4.1 Animism

The closely related concept of **Animism** is as defined by Piaget [Piaget, 1933] and most following psychologist the "Attribution of life to the nonliving". Similarly, religionist and anthropologists look at animism as the attribution of soul or spirits to things. The animistic worldview originates from the same source as anthropomorphism: facing uncertainty, we all look first for what matters most - living things and especially humans - and implicitly assume these as a default. Animism is less complex than anthropomorphism, and with exception of ethnographers and anthropologists it was mostly of interest for the domain of child psychology [Mead, 1932] [Bullock, 1985]. Although this phenomenon attracted less attention in cognitive sciences, the transition to anthropomorphism is rather seamless, when it comes to the attribution of life-like characteristics to inanimate objects.

From a definition point of view, animism is applied as long as only animalistic attributes are applied that do not include human traits such as etiquette, gender specifics or emotions that go beyond primal desires.

2.4.2 Anthropomorphism in Human-Computer-Interfaces

Taking a look at contemporary HCI on the background of the work summarised in the previous section, we will find diverse research activities that have shown how users sometimes interact with desktop PC applications as they were interacting with people [Lee and Nass, 2003] [Nass et al., 1999] [Nass et al., 1997]. People apply well-learned conventional social schemas such as gender stereotypes and norms such as reciprocity and etiquette when they respond to the interactive system [Nass, 2004]. There is also evidence that users simply apply stereotypes and heuristics and enact social habits [J.J. Bargh and Burrows, 1996]. It can also be observed in the technical language of computing how humans apply anthropomorphism as a framework to comprehend technology: A computer has a *memory*, it *searches* for things, it is programmed by programming *languages* and can be infected by a *virus*. This is also supported by the the CASA paradigm (**C**omputers **A**re **S**ocial **A**ctors) [Nass et al., 1994]. Many users state that they interact socially with their computer, both verbally and physically and that they often regard it as a friend, enemy or both. Of the 53% of participants who admitted talking to their computer kindly, 90% also admitted that they speak to their car. Of the 47% who said that they did not talk to their computer kindly, only 68% claimed that they talk to their car. This seems to indicate that some users might be predisposed to talking to certain objects, whereas others are not.

The corresponding user interface counterpart for animism is termed *animalistic* or *zoomorphic* user interface (see e.g. [Fishkin et al., 2000a]), and is referred to when simplified life-like entities are employed that do not exhibit human capabilities such as spoken natural language.

Human computer interaction designers have already attempted to build upon the mechanisms of anthropomorphism in a variety of applications, for example life-like characters as for instance described by André et al. [André et al., 1998] or Cassel et al. [Cassell et al., 1998].

Life-like characters, also referred to as Virtual Characters, Avatars, Embodied Conversational Agents or Anthropomorphic Agents, are graphically represented believable characters, which can be represented in a realistic or also in a cartoon-style manner. A lot of research energy is spend for examining and increasing the perceived believability and life-likeliness of such characters, which depends, according to Hayes, on whether a character seems conversational, intelligent, individual, social, emphatic, variable, and coherent [Hayes-Roth and Doyle, 1998]. Such characters are increasingly employed for tasks that are originally performed by humans, such as tutors, trainers or sales persona and aim at applying an anthropomorphic interaction metaphor in order to increase the otherwise poor communicative capabilities of computational devices. Robots due to their inherent physicality and ability to move, sense and react, exhibit an obvious instance of machines that also have the potential to be exposed to an anthropomorphic stance by users. It was for instance shown that *"people developed a predictable mental model of the robot's knowledge in an entirely different domain (tourist landmarks). Just as they do for people and animals, they made inferences about the robot's internal knowledge state and extrapolated to predict its competencies [Sara et al., 2005]."*

Further examples emerge from contemporary pop culture and can be found in movies and literature that envision anthropomorphised interaction with computer systems, e.g. the spaceship main computer in the movie "2001 Space Odyssey" (see [Schmitz et al., 2008b] for an overview). An abundance of commercial Product Design examples and marketing concepts follow an anthropomorphic scheme (e.g. M&Ms, Mr. Clean or J.P Gaultier perfume).

2.4.3 Software Agents

Another prominent example of transferring anthropomorphic concepts to user interfaces are software agents [Maes, 1994], which were introduced as a novel interface paradigm constituting an approach to model software behaviour and algorithms as self contained actors. This concept was not intended as an anthropomorphic approach as discussed in this chapter, instead it was motivated from a system design point of view as an alternative interaction style to direct manipulation. The discussed anthropomorphic features are not necessarily given in software agents, but fact that they constitute computer software appearing as independent actors instead of tools is an essential attribute of both concepts. One major differentiator is that users delegate tasks to the agent-based software, which autonomously works on the given problem and presents the result afterwards.

According to Woolridge and Jennings, there is no widely acknowledged strict definition of software agents, but it has been commonly approved that agents are computing systems with some sort of autonomy. Nevertheless, they attempt to mark the term agents as systems that are situated in an environment, and that are capable of autonomous actions in this environment in order to accomplish certain objectives [Wooldridge and Jennings, 1995]. Similarly, Russel and Norvig describe agents as *"anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors."* [Russell and Norvig, 2009] - which would also subsume TASOs that employ sensing and actuation capabilities (see also Figure 2.2).

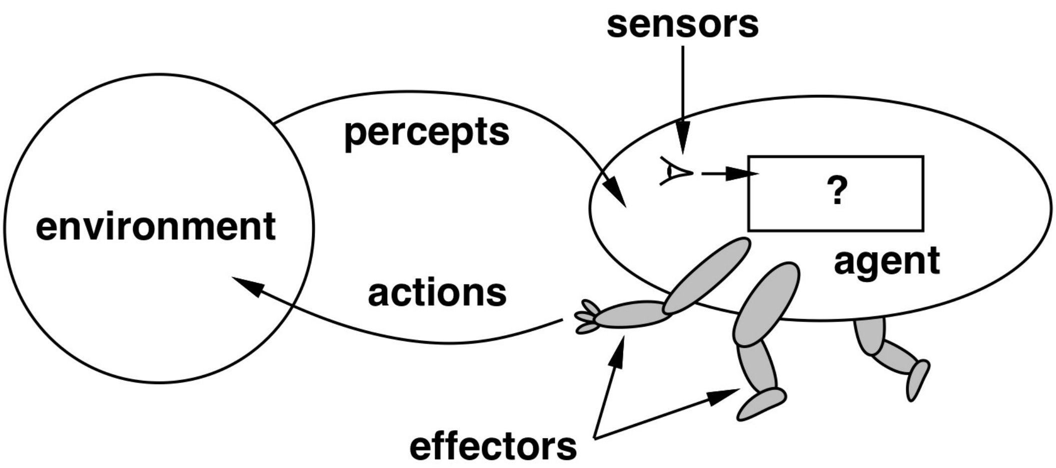


***Figure 2.2**: Russel and Norvig's model of an agent (source: [Russell and Norvig, 2009])*

Woolridge and Jennings further introduce the ”Weak Notion of Agency”, a general description of required attributes of an entity to be called an agent [Wooldridge and Jennings, 1995]:

Autonomy: Agents are in control of their internal states and actions, such that they can operate without explicit intervention by humans.

Social ability: Agents interact with other agents or also humans in order to e.g. negotiate or cooperate.

Reactivity: Agents perceive their environment, which might be facilitated through e.g. sensors, and respond to changes.

Pro-activeness: Agents are not limited to simply react but are also able to take the initiative for goal-directed behaviour.

A stronger notion of agents, e.g. discussed by Shoham, suggests the attribution of intentional properties to agents, also called *mentalistic* capabilities [Shoham, 1993]. Such capabilities include for instance the concepts of belief, desire, decision, intention, or obligation. This notion makes use of well established folk psychological theories, which are applied to explain and predict human behaviour (see also Section 2.4.4).

2.4.4 Layers of Anthropomorphism

A study of recent anthropomorphic user interfaces on the background of psychological and anthropological research has shown that such systems and their approach to exploit anthropomorphic reactions in users is multilayered and involves several processes on rather independent levels. They argue that anthropomorphism in human computer interfaces includes a variety of phenomena that are not yet thoroughly understood, and as a first step categorised anthropomorphic phenomena in six layers [Persson et al., 2000] [Persson et al., 2002] :

- *Primitive categorisation* is the most basic level of anthropomorphic thinking and involves the general impression of ”life” triggered e.g. through movement, visual properties such as faces or extremities or through the presence of voice or other animalistic sounds. When primitive categorisation is in place, people start projecting other aspects on a given living thing.

- *Primitive psychology* describes individual everyday knowledge about basic needs and drives, such as hunger and pain, life preservation, sexual drives or sensations. For example, hunger will disappear after eating. Looking at the definitions above this level of anthropomorphism can also be seen as the layer of animism or zoomorphism, as a pre-stage of full anthropomorphism, although the authors do not differentiate between the concepts of anthropomorphism and animism at this point.

- *Folk psychology* encompasses more complex (folk-)theories of the human mind, explaining behaviour in terms of inner states that relate to perceptions, beliefs, goals, intentions and actions. For instance, goals motivate intentions and beliefs constraint actions. Furthermore folk psychology involves the attribution of *emotions*, which is further defined by the appraisal theory, which tries to describe how people appraise and ascribe emotion to others (e.g. [Omdahl, 1995]). The central idea is that emotions result from our evaluations of events or situations, causing different reactions in different people. Therefore each situation causes an emotional response in a person based on the his or her appraisal of the event. Attributing emotions to other people always involves such appraisals.

- *Traits* are summaries of one's impression of a person, and form another layer of anthropomorphism. Traits such as *idealistic*, *thoughtless*, *stupid* are common personality stereotypes and describe simplified everyday-versions of formal concepts of personality psychology, such as the prominent five-factor-model of personality [McRae and John, 1992]. Another important difference of such traits compared to the folk psychological level is the endurance and stability of these features. In this regard, this definition is equivalent to a classification of emotions and personality as in [Gebhard, 2009], where durability is also the main feature for differentiation. It further appears that traits seem to be shortcut terms for related processes on the level of folk psychology: For instance an *impulsive* person disregards *thinking* and *beliefs* as constraints for *actions* and instead acts mainly on *desires*.

- *Social roles* are applied in everyday encounters to explain situations and to provide behavioural patterns through a social schema. Several social schemas were proposed by psychologists and sociologists, constituting fundamental dimensions in socio-cultural environments. *Occupancy roles* provide us with normative expectations about e.g. doctors, waiters or bus drivers and possibly about their goals, beliefs, emotions etc. *Family role schemas* regard father, mother, brother etc. and contain expectations about their interaction on a daily basis. *Social stereotypes* constitute a mix of assumptions about other people and can be conceived as a set of traits associated to a group of people. For instance in our culture, women are considered to be emotional and empathic, while doctors and scientists are rational and emotionally neutral. In contrast to other social schemas these stereotypes are often mixed with an emotional and moral verdict about this category of people. Furthermore, expectations and associations for stereotypes are often linked with salient visual features, clearly identifiably in first encounters, including skin and hair colour, body size, gender, age etc. With increased familiarity with a person, visual cues become less important.
- *Emotional anthropomorphism* is the last level of anthropomorphism and - in contrast to the other levels - does not involve cognitive processes but an emotional dimension that evolves over time. The term emotional bonding coins the phenomenon that people do not only try to understand the behaviour of other individuals (by applying folk psychology, traits and/or social roles) but also take an emotional stance towards another person, by for instance aligning with a character or person or applying a process of allegiance. It has to be noted that these processes including long-term emotional bonding processes like friendship and love, and their complex and dynamic interactions are not yet fully understood by neither psychology nor anthropology.

Examining this categorisation more closely, we will notice that each of these layers incorporates an increasingly complex model of an anthropomorphised subject building on top of previous layers. Three dimensions can be identified which differentiate one layer from another, all of which are increasing with each layer:

- Complexity of the anthropomorphic entity's personality
- Duration required to build up the intended anthropomorphic phenomenon
- Relevance of social context

Primitive categorisation takes place instantly, potentially at first sight, and only suggests the mere existence of life-likeliness. Primitive psychology just needs an experience of action or reaction in a certain context that can be evaluated, which can basically happen within seconds or minutes. Folk psychological phenomena involve more complex explanation processes of observed behaviour, while traits and social roles regard compound identities that consist of sets of personality factors. Social roles furthermore build on social context and inter-individual dependencies combined with personality types. The emotional anthropomorphism particularly requires longer time periods and mainly relies on emotional relationship instead of individual attributes, while the latter are also required as a pre-condition.

2.5 Synopsis

In this chapter we outlined the current research context of this work and described essential concepts on which the dissertation builds on. We have summarised recent research developments in the area of UbiComp and discussed the the current state and recent transitions of interaction design concepts. UbiComp became research reality from vision due to developments on various frontiers, primarily the miniaturisation and mass production of computing and wireless communication components, sensors and actuators. It is further supported by advancements in processing and interpreting complex sensory information, such as in visual computing, speech recognition, affective computing, gesture recognition or sensor fusion. We have seen that several research trends approach the field on different abstraction layers, for instance Ambient Intelligence with a focus on user friendliness and services, or Pervasive Computing, which concentrates on infrastructures and middleware. Research on instrumented environments and smart objects can be seen in this context as *enabling technologies* for UbiComp scenarios, applications and user interfaces. Particularly for the technical prototyping part of this work, it is important to note that smart objects as defined in this chapter can be infrastructure independent entities but also be part of an instrumented environment and access services (e.g. sensors) of the environment in order to employ the intended functionality and interfaces. We have pointed out that various current projects build on smart objects as enabling technologies, for example the SemProM project, which employs product diaries for value-added services. In fact, these diaries can already provoke an anthropomorphic stance, because they communicate that an object has undergone *experiences* that affect the current state of the object. Thus, sensory perception builds up object memories and memories define coherent current states, which potentially supports the believability of life-like objects.

Typical UbiComp scenarios involve everyday environments with an abundance of smart objects that a user has to cope with in a preferably intuitive and enjoyable manner. In such environments, users might be exposed to as many different services as there are objects, each of which potentially with its own interface. The TASO concept constitutes an interface paradigm that attempts to relieve users from the burden of memorising and internalising a variety of independent interaction principles by providing a holistic paradigm that leads to life-like smart objects in many variations.

We have discussed facets of evaluating qualities of interaction. The transition from a comparably pragmatic view on usability towards the inclusion of hedonic attributes is described in Section 2.2 and opens new venues to develop and evaluate the qualities of novel interfaces, which might not exclusively aim at highly efficient processes but also at emotional and engaging user experiences. We believe that the concept of TASOs has a high potential in satisfying the stated quality criteria in UbiComp scenarios. The concepts proposed later in this work address a series of criteria that have been identified as key factors for establishing more joy of use.

We have reviewed tangible interaction and some crucial concepts behind this interface paradigm. The idea of TUIs comprises a specific concept for interface design, irrespective of technologies and implementations. UbiComp can provide enabling technologies for implementing and realising TUIs, and in return, TUIs can be integral parts of UbiComp scenarios and provide interaction concepts that potentially apply well to many use cases. We have discussed the impact of some fundamental constructs in the realm of tangible interaction, including affordances and a taxonomy based on *embodiment* and *metaphor* and we will see later how the design of TASOs can draw from these concepts. We have also seen how conceptual research such as unifying classifications attempt to organise and relate ongoing research in the field of HCI in UbiComp scenarios and how the concept of combining tangible interaction, anthropomorphism and smart objects embeds into current research context and discloses new territories.

Anthropomorphic interfaces and TUIs share the common goal to exploit established human skills and knowledge acquired in everyday life. We have also seen that animism as a special case of anthropomorphism describes more simple phenomena which we applied to construct an animalistic class of smart objects (see Section 4.3.1) in this work. We have introduced software agents and shown how they employ a certain anthropomorphic paradigm as system design strategy.

Furthermore, we introduced several layers of anthropomorphism based on anthropological and psychological studies and observations. We will adapt and apply these observations for building distinct classes of anthropomorphic objects in Section 4.3, providing alternatives for different scenarios, depending on their requirements and characteristics.

3 RELATED WORK

The previous chapter provided a clarification of terms and a solid background in research activities relevant for this work, primarily interaction design in the era of UbiComp, tangible interaction and anthropomorphic user interfaces. In this chapter we give an overview on state-of-the art in these research areas by presenting a selection of projects that are relevant for this dissertation.

This chapter is divided into four sections, analogous to the previous one. We start with reviewing smart object interfaces that don't adhere to any particular interface paradigm, followed by work in the field of TUIs. We will continue with user interfaces that involve anthropomorphic and animalistic concepts, ranging from embodied conversational agents to social robots and toys. Finally, we conclude with a discussion of the reviewed projects, their advantages, results and shortcomings related to TASOs.

3.1 Smart Object User Interfaces

Interaction paradigms for smart object systems are few due to the rather young age of this discipline. In this section we will introduce various kinds of user interfaces for smart objects, which have emerged in recent years. The spectrum includes implicit interaction, direct manipulation, remote interfaces and mobile phone mediated interaction styles. In contrast to the interface paradigms introduced in other sections of this chapter, the approaches of the smart object interfaces introduced in the following are often technology inspired and results of an opportunistic approach to realise interfaces to objects that are already augmented with technology and thus offer certain sensing capabilities for object manipulation.

3.1.1 Implicit Interaction With Smart Objects

We will use the term *implicit interaction* to subsume interaction paradigms that involve the use of situational context information, such as the location and state of smart objects and users. Implicit system inputs are therefore actions of the user, which are not perceived as such, instead they are real world actions with an independent purpose. Albrecht Schmidt has coined this term in 2000 and proposed the use of context in order to improve user interfaces [Schmidt, 2000]. The application ContextNotePad was introduced to exemplify this idea and included context adaptive features, such as hiding the display content, when the device was not in use and other people are nearby.

This approach is further picked up in [Schmitz et al., 2007a] and conceptually applied to the shopping domain: In this scenario a user in a supermarket might take a product out of the shelf, turning the product in order to read what is written on the outer packing, gathering information about the product, instructions to use, additionally needed accessories or in case of groceries the nutrition facts and serving suggestions. Actions like picking up/putting back products from/to the shelf or shopping basket can for instance be detected using RFID technology. Interactions with the products outer packing can be monitored using sensor boards, which are able to provide sensing capabilities e.g. for light, temperature, sound and acceleration. Such observed interactions could trigger events in an UbiComp environment the user is not aware of. In the case, where the user in our scenario puts the product back to the shelf, it would for example be possible for such a value-added service to compile a personalised product catalogue, that is somehow (e.g. via e-mail, bluetooth transmission, or printed version) handed over to the user at the end of his shopping process. Or the system might automatically check the nutrition facts against the user's allergy profile and warn him, whether the product contains ingredients that might cause allergic reactions. As these examples illustrate, the interactions are typically unnoticed and the design of the interface does not intend a designated explicit interaction - as opposed to the concept TUIs.

3.1.1.1 The Smart Shopping Assistant

In this fashion Schneider has developed a prototype of a smart shopping assistant [Schneider, 2004], which reacts on certain actions of the user: Taking a product out of the shelf and placing it into the shopping cart. Upon the first action the system will display general product information on a display attached to the cart and additionally compare two products if another item is taken out of the shelf at the same time as shown in Figure 3.1. The second action will cause the shopping assistant to evaluate the current set of products inside the cart in order to infer, whether the user is going to prepare a certain recipe. The system will provide a list of recipes, which require products that have already been placed into the cart and orders them according to the probability that the the user is going to prepare this dish.

***Figure 3.1**: The Smart Shopping Assistant in use*

Clicking on one of the recipes will display the list of required items, indicating which of them are still missing. This example of implicit user interaction also shows the potentially blurred boundaries to explicit interaction: As soon as the user believes that the interaction pattern of the system has been completely induced, she or he can attempt to exploit its behaviour to actively trigger desired responses, which might turn an implicit interaction style into an at least partially explicit one. For example, a user might place a couple of products into the cart in order to get a certain recipe displayed on the list, just to be able to check the required preparation time.

3.1.1.2 The MediaCup

Some of the first prototypes illustrating the potentials of implicit interaction based on context gathered by augmenting everyday artefacts with sensing and communication abilities was developed within the MediaCup project [Beigl et al., 2001].

***Figure 3.2**: The augmented MediaCup (source: [Beigl et al., 2001])*

The MediaCup itself is still a prominent instance of these efforts: An ordinary coffee mug with hardware embedded into the base, containing both hardware and software for sensing, processing and communicating the state of the cup (see Figure 3.2). A coffee cup was chosen as the central item, because it constitutes a typical everyday object that is frequently used while remaining in the background of the user's attention. While a focus was on communication infrastructure and technical issues, several application scenarios developed over time. The most obvious step was to detect whether the cup is filled with hot coffee or not and to let the coffee machine start brewing replenishments if appropriate. Further, so called Smart Doorplates, displays with dynamic contents attached to the outside of office doors, would be able to detect whether there is a meeting situation and to show a ”meeting” warning to potential visitors. Such a meeting situation is basically assumed when more than one cup are present in an office or meeting room and being used.

The exactly same approach is taken by the Sentient Artefacts project [Kawsar et al., 2005], although a the name of their concept does not pick-up the prevalent terms. Again, everyday objects are augmented with sensors to enable unobtrusive and implicit value-added services to the user. The smart artefacts here are also seen as aggregator of context information, including the state of the object and user activities. Several applications for home environments were developed, for example the AwareMirror [Fujinami et al., 2005], a mirror with display capabilities installed in a washroom, showing various information such as the schedule of the day or weather forecasts in addition to the person's reflection. A toothbrush is augmented with two-axis accelerometer and actually used as sensor that allows to infer the use of the bathroom, which initiates the mirror interaction.

3.1.1.3 Smart Mats, Tags and Clips

Preceding the development of our prototyping environment (see Section 5.2) we have constructed a series of smart objects by integrating different kinds of sensor technologies partly into existing objects and partly into devices specifically created for this purpose. The experiences of this project were incorporated in the development of our final prototyping testbed and also helped us to develop industrial design guidelines presented in Sections 4.1.5 and 4.2. In the following we will describe selected products that have been created in the context of an interdisciplinary project involving students of computer science and product design, who were given the task to realise fully functional TUIs under our technological and conceptual guidance.

Figure 3.3*: The interactive beer mat.*

In this project we hoped to bridge these two worlds of design and technology and to achieve a design process, which is informed from both sides. While the product design students could still do their original work of finding and developing forms and shapes, they received permanent feedback about how well their designs were suited to being turned into a functional prototype with the technology at hand. They learned to pay attention to this aspect early in their designs. Conversely, the computer science students learned a lot about the restrictions implied by mechanical design and were forced to be more creative in their choice of technology. We expected that the emerging designs would gain substantially from this constant dialog between the free flow of ideas and the necessity to produce a fully functional prototype in the end.

The first product we want to introduce is the **interactive beermat**, which is also an example for interfaces that deliberately provide both implicit and explicit interaction modes [Butz and Schmitz, 2005]. The device is basically a reusable beer coaster made of plastic with integrated weight and acceleration sensors (see Figure 3.3) .

The implicit part is realised with the weight sensor, which detects whether a glass on top of the coaster is empty and in that case sends that information to the desktop computer, which would theoretically notify the waiter. The beer mat is further also intended to support entertainment activities in pubs, by using acceleration sensors to sense orientation in space, which can be used for voting processes. While simply raising the glass is usually associated with a positive reaction, a negative vote can be given by raising the glass and explicitly turning the mat upside down. Voting can be used in song contests to determine the vote of the audience or in karaoke bars to give immediate feedback to the singer already during performance. In sports bars, the decisions of the referee on the big TV screen are often cause of discussion, and the collective voting can here convey the average opinion of the local pub crowd, thus creating an additional level of interactivity for watching sports in a group.

***Figure 3.4**: The smart security tag.*

The **smart security tag** is an extension of conventional security tags that are attached to clothes and also other products, in order to trigger alarms when carried through the exit [Kaiser, 2004]. The basic idea is to have such conventional tags extended by the means of generating simple visual signals that indicate whether that particular piece suits a given customer profile or not. This profile could contain information about age, gender, size and price range, depending on the final scenario and available information. The developed prototype applies a fixed set of selected attributes and matches them against available items. A subset of tags which is associated to the items that are potentially interesting for the customer, will glow as long as the user is present and ”logged in” at the shelf, which was realised by an RFID-based recognition of a personal item, such as a customer card (see Figure 3.4).

The smart tag employs a active RFID tag extended with an additional LED, communicating with the *i-Port III* RFID reader. The communication of between the RFID tags and the reader is divided into two steps:First, the reader scans on the connected antennas of the i-Port to determinate which tags are within range. Second, a blink command is sent to the corresponding RFID tags. The first design concepts involved *Smart-Its* sensors (predecessors of the pPart Particles, see Table 5.1.1), but since the final version did not require additional sensors, the technology requirements have changed such that the size of the device could also be reduced.

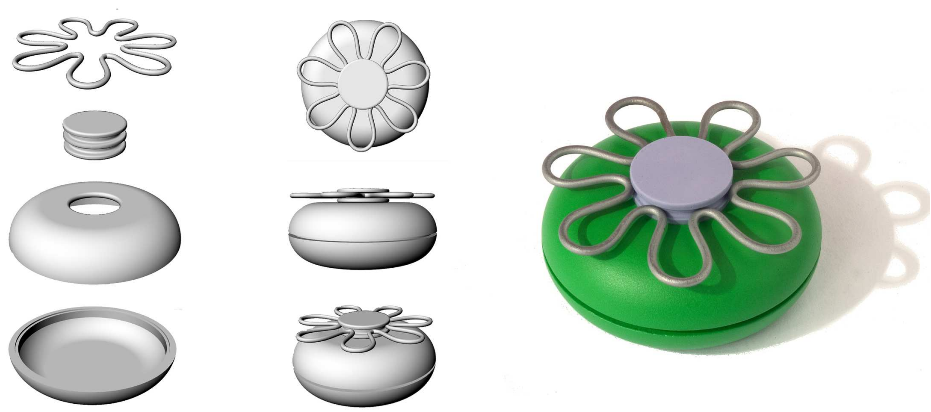


Figure 3.5*: The smart clip.*

The **smart clip** is a brooch that supports gymnastic lessons by tracking movements with acceleration sensors and sending the data to the main application which displays feedback on a nearby screen [Berwanger, 2004].

The brooch itself has to be fixed to a particular position at the user in order to facilitate the recognition of the movements (see Figure 3.5). The requirements of the case therefore include the possibility to easily (re-)attach it to clothes and certainly to hold the sensor device and to provide breakouts for a recharger and a power switch. At the same time the size has to be minimised to make it as unobtrusive as possible.

The main software is located on a common PC and mainly analyses incoming acceleration data in order to detect whether the movements were performed in the intended manner. The actual feedback was performed by displaying video clips that contain either affirmations or corrections of a few selected movements.

3.1.2 Mobile Interaction With Smart Objects

The core idea of the the Perci (PERvasive ServiCe Interaction) project, which was funded by the NTT DoCoMo Euro-Labs, is the use of mobile phones as mediators for interaction with smart objects [Enrico Rukzio, 2008] [Enrico Rukzio, 2007] [Broll et al., 2009]. This is motivated by the rapid development and use of mobile phones with integrated built-in cameras, GPS receivers, Wifi and Bluetooth hardware, which are the enabling interactions with objects in the real world. Further, technologies like RFID, Near Field Communication (NFC) or visual marker recognition are can already be found in some newer models, providing additional means for object recognition. Smart objects have to be augmented such that they can be sensed by a mobile device. Application examples for the technical infrastructure developed in this project are advertisement posters augmented by visual markers, machines (e.g. a printer) augmented by RFID/NFC tags and public displays, which are augmented by Bluetooth-based services accessible via the mobile device (see Figure 3.6).

Figure 3.6: *Touching and pointing for interaction with smart objects (source: [Broll et al., 2009])*

Thus, a smart object from their understanding can be a location, a physical object augmented visually or on RFID basis, another mobile device or a computer providing services via Bluetooth. The technical implementation suggests at the same time the interaction style. The authors differentiate between touching, pointing and scanning. Touching is facilitated and required by a RFID/NFC implementation, since the recognition of passive RFIDs works only for a very short range. Pointing takes place when visual recognition is required using the build-in camera of the mobile device. An alternative solution for the pointing interaction has been developed with a laser pointer attached to the phone and a light sensor mounted on the smart object. Scanning is the last of the three proposed interaction styles and actually describes the phone's search over bluetooth for smart objects that are also bluetooth enabled and that the system can connect to. The applications realised in this way basically translate IDs of smart objects to pointers that direct the mobile device to web services, for example to an online shop for movie tickets for the film advertised on a poster.

This approach separates the interface from the smart object, which makes the physical appearance of it almost irrelevant. The interaction modes are shaped by technical restrictions and basically allow access to services attached to physical objects. The advantage they draw on is the pervasiveness of mobile phones and the fact that the users are familiar with their personal device.

3.1.3 Commercial Products for Smart Object Creation

The success of RFID in research and industrial applications and the emergence of the internet of things has recently also lead to start-up companies that provide the means to configure smart object applications for end consumers. The company Touchatag[1] is one of them and has emanated from Bell laboratories, the research division of Alcatel-Lucent. Touchatag basically sells RFID equipment, that is sets of readers and tags and provides an easy to use system to associate tags with configurable functions, such that for instance a predefined email will be sent when the associated tag is detected by the reader (see Figure 3.7). Users can now stick RFID tags to suitable everyday objects like cups, wallets or books etc., place the reader for instance under the kitchen table and choose a response of the system that is triggered when the object is in range of the reader. The company also provides an open marketplace for such services and documented libraries for developers, such that the community can contribute to the functionality of the system.

The french company Violet[2] was founded in 2003 and offers a product with a set of features very similar to the Touchatag: Sets of RFID readers and tags, namely Mir:ror and Ztamp:s, and simple means to associate actions to tags, thus adding new functions to everyday objects. In addition to that Violet also developed the Nabaztag, a configurable device iconically resembling a rabbit and containing a variety of sensors and actuators (see Figure 3.7). This includes a Wifi card for wireless internet access, a loudspeaker, LEDs, microphone, RFID reader and motorised ears that can rotate and detect when they are moved by the user. The Nabaztag's behaviour is configurable over a web interface and further programmable via an API. It can blink, move its ears, react on spoken commands through voice recognition, read out loud texts utilising a text-to-speech module, detect RFID tagged objects and react on incoming mails, RSS feeds and various other web resources. The variety of applications is certainly more diverse as with simple RFID recognition and thus, the set of services is extended by the community and continuously growing.

Although the Nabaztag is not an everyday smart object but instead a specialised interactive device basically for delivering web-based services, which is also supposed to blend into the environment as a decorative item, it represents a mature product in the periphery of the smart object domain that has made it to the end consumer and market. The design of the shape, the text-to-speech functionality and the movement of the ears furthermore provide a mixed animalistic/anthropomorphic impression. While the interaction style itself, the

[1] `http://www.touchatag.com/`, last visited June 30, 2010
[2] `http://www.violet.net/`, last visited June 30, 2010

***Figure 3.7**: Left: The Nabaztag - Right: A Touchatag set incl. RFID reader and tags (sources: http://www.violet.net/ and http://www.touchatag.com/)*

configuration and programming all happen on traditional channels and break with the animalistic/anthropomorphic approach.

3.1.4 Writing and Talking to Everyday Objects

Recent projects address the technical challenges of creating responsive artefacts that react to written and also spoken messages. Choi et al. suggest to interface smart devices with common instant messaging clients as if each device were a buddy to chat with on text basis [Choi and Yoo, 2008]. This includes messages that initiate communication (by typing *"hello"*) and to give commands (*"please print document xxx"*). Also feedback and system status are received as messages in the same way. On the one hand this is a rather simple technical solution to deliver commands with existing software to certain machines, but on the other hand it invokes a awareness of things as dialogue partners and somehow responsive entities, which were not perceived as such beforehand. Certainly it is purely virtual, since it only works on common personal computers or other machines that support chat clients and therefore the physical aspects of smart objects are totally excluded.

While Wasinger et al. were the first to examine the effects of applying spoken dialogues to objects (see Section 3.2.2.2), others have picked up this idea to motivate their work on technical realisations of such interfaces. For instance, Lombriser et al. have realised speech as a modality for smart object interaction [Lombriser et al., 2009], arguing that speech has the potential to provide naturalistic and intuitive interaction. The authors have examined the use of speech synthesis on sensor nodes that may be integrated into smart objects. For this purpose the so-called Wireless Voice Node has been developed, which is a small, wireless sensor node with the ability to generate voice output with a voice synthesiser and integrated text-to-speech processor or an audio system that plays pre-recorded sound files stored on the board. In addition to that the developers added a transceiver for wireless communication, a 3-axis accelerometer to detect gravity and movements performed with the device, an air pressure and temperature sensor to be able to detect height and temperature changes and finally a light sensor calibrated for indoor usage. Because power consumption is a particular issue for wireless computing it was an interesting point that voice output is a very expensive process, accounting for 90% of the power consumption of the sensor board when turned on.

The first example use case of their project is the talking doll, a puppet for children with a programmed voice node integrated into the head (as shown in Figure 3.8). The aims of this prototype were to enhance the playing experience of children by motivating them to use other, wirelessly connected smart toys. Thus, the doll will speak context dependent sentences when children interact with those toys.

Figure 3.8: *Integration of the Wireless Voice Node into the speaking doll: a) the Wireless Voice Node, b) loudspeaker, c) battery, and d) a milk bottle as smart toy (source: [Lombriser et al., 2009])*

From time to time the doll will ask for another augmented toy in the environment. All smart toys send out their identifier strings (e.g. ”milk” for the milk bottle), which can be read out loud by the doll when it asks to play with it (e.g. ”I want drink *milk*”). Utilising the on-board sensors the doll and the toys can recognise motions and send out further strings in response to that, which can further be interpreted by the doll. In the example of the milk bottle, the bottle itself would recognise whether it is tilted as if it were used for drinking and send out *drinking* to the doll, which in turn could render drinking sounds that were stored on the board. Action sentences and sounds are only played if the smart object seems to be in immediate proximity of the doll, which is determined by evaluating the received signal strength. Details about the algorithm and implementation can be found in [Lombriser et al., 2009] and [Brunette et al., 2003]. Since the doll does not store any information about potential smart toys it interacts with, the system can also integrate new toys that were not known at design time. The prototypical implementation of the smart doll instance is mainly a proof-of-concept of the wireless sensor and speech synthesis node and has not been further investigated from the user interface perspective.

3.1.5 Summary

The first interface examples in this section demonstrate a major benefit of integrated miniaturised technology: Smart objects can utilise sensors not only to provide additional services but also to detect user input. For example, acceleration sensors can detect both whether an object has been dropped and whether a person is holding it in a certain way. The advantage of implicit interaction is on the one hand to realise unintrusive services to the user without requiring any sort of explicit user input, such that the user is not required to learn any means of interaction or to adapt his or her natural activities. But on the other hand an explicit interaction style potentially provides more control to the user in certain situations. The interactive beer mat exemplifies how explicit and implicit interaction can be combined in one smart object. Important for this work is to see how user input is recognised by the introduced projects. A basic and common method used by the Smart Shopping Assistant is to use RFID technology to infer whether a user is holding a product in her hands, which is done by simply installing a reader in shelves such that the user can be assumed to hold a particular product when the absence of this object is reported to the system. The smart object in this case merely holds its identification tag while the environment provides the remaining technology, including the output device. The Nabaztag also hosts all sensors and even output, computing and communication technology. Obviously, as a commercial product, specialised infrastructure to support smart object interaction cannot be assumed at the consumers home yet. Although the anthropomorphic approach is only superficial, it is interesting to note that an animal-like appearance was chosen, probably to develop a consistent and pleasing product impression for a device that offers both reactive and pro-active functionality.

It seems apparent that several examples in this section were developed in an technology-inspired fashion, which means that the possibilities of available sensor platforms were explored for interactive purposes without being motivated from a perspective that focuses on the application itself or even the interaction design.

3.2 Projects Involving Tangible Interaction

Typical tangible user interfaces focus on (sometimes two-handed) manipulation of physical artefacts, which are designed for the particular system and input method. Besides such projects we will also present work that integrates tangible input into multimodal systems and prototypes that combine TUIs with life-like characters.

3.2.1 Tangible User Interfaces

We will begin our review on related tangible interaction research with classical TUIs, which focus on the tangibility of the interfaces and their physical manipulation. The selection includes aged but highly influential work as well as state-of-the-art projects covering the various research directions of the TUI realm.

3.2.1.1 musicBottles

The Tangible Media Group of the MIT Media Labs has created a series of pioneering prototypes, such as the musicBottles [Ishii et al., 1999] [Ishii et al., 2001]. The musicBottles interface employs bottles as containers and controls for digital content and consists of an instrumented table and corked bottles containing various acoustic media, e.g. musical instruments (see Figure 3.9).

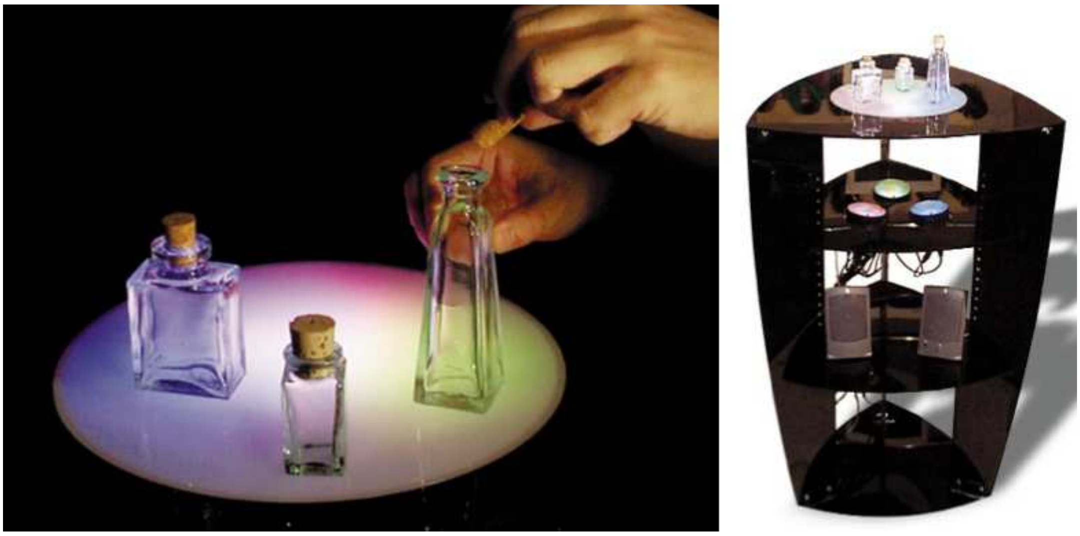


Figure 3.9*: The musicBottles (source: [Ishii et al., 1999])*

The system recognises each bottle that is placed onto the table through RFID tags and lights up the stage to show that the bottle has been identified. Opening a bottle is also detected and "releases" the contents by rendering the corresponding sounds and back-projecting coloured light patterns on the table, reflecting changes in pitch and volume of the sounds in realtime. The user is therefore able to orchestrate the acoustic and visual experience by applying the interaction metaphor of opening physical bottles. The initial prototype employed the musical instruments violin, cello and piano in Edouard Lalo's Piano Trio in C Minor, but the group has also experimented with other applications including weather reports, poems, and stories.

3.2.1.2 A Workbench for Urban Planning and Design

The URP-System [Underkoffler and Ishii, 1999] supports architects and town planners in certain planning tasks: Users position shape models of buildings on a planning table, while illumination and shadowing conditions are computed and projected in realtime onto the surface (see Figure 3.10). Additional objects beside the building models function as tools to manipulate parameters of the system or objects. For example, a magic wand can be used to change the material of a wall to stone or glass by touching it. The reflections of walls made of glass will be included into the computation of the lighting conditions. Another example is the ruler, which can be utilised by touching two points on the table to let the system compute the distance and display the value on the surface. The clock is another important tool to set the time at which the illumination and shadowing should be computed. This allows for instance to observe whether a courtyard gets enough light during the course of a day. The flow of wind is also computed and blended into the projection, visualising phenomena such as suctions caused by certain building configurations, or swirls in between buildings. The winds are represented by short lines moving in the direction of the winds, supplemented by the speed index.

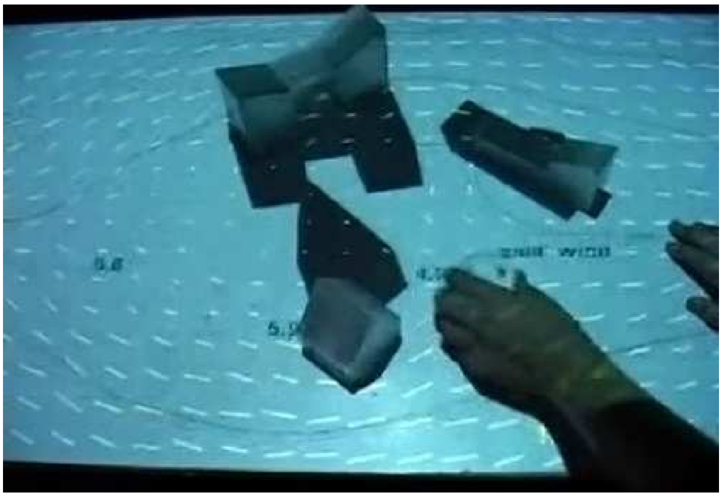


Figure 3.10*: The URP system (source: [Underkoffler and Ishii, 1999])*

The original system was based on a visual recognition approach, although later versions of the table employed physical sensors integrated into the table and models. The realtime response to user manipulations is already a significant improvement over common simulation applications in the city planning domain, enhancing the typical iterative planning process towards the desired result. All meaningful parameters can be physically and intuitively manipulated, while the dynamic content is delivered digitally through the projection. These features lead to another important benefit of the overall system noted by users: The possibility to solve problems in a very playful and iterative manner.

In later generations the URP table evolved to a more sophisticated application supporting more realistic settings and features, which filled the surface with heavy clutter of objects representing new functions and buildings [Ishii et al., 2002]. In consequence the developers decided to keep certain functions purely digital, accessible with a traditional computer mouse.

3.2.1.3 MediaBlocks

The MediaBlocks system [Ullmer et al., 1998] serves as a tool for the creation, management and manipulation of media, which are dynamically associated with wooden blocks. These blocks are digitally tagged and can be recognised by certain surfaces, for example one that belongs to a video camera, display or printer. Following the idea of the marble answering machine, these blocks are used as physical containers of data that can now be easily transferred between different devices. A central component is the so called MediaSequencer, a tangible interface for mixing and manipulating media (see Figure 3.11) .

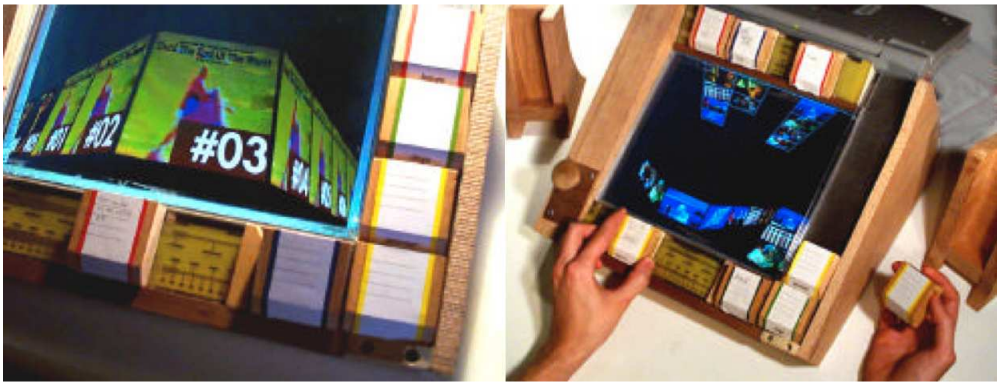


Figure 3.11: *The Mediablocks sequencer (source: [Underkoffler and Ishii, 1999])*

The shape of the MediaSequencer reminds of a picture frame, with a screen in the centre and one rack at each side equipped with sensors to recognise MediaBlocks placed on it. The two main components are the position rack (bottom) and the sequence rack (top). When a mediaBlock is placed in the position rack, the display shows its contents as a foldable series of images, as shown in Figure 3.11 in the right. The focus of the view is controlled by the horizontal position of the mediaBlock on the position rack: Moving it to the edge on the left side will focus on the first element and, accordingly, the most right position corresponds to the last element of the block's media content. These functions can be used to browse the contents of a mediaBlock and to select an individual media element for copying between mediaBlocks. For the latter process, the destination mediaBlock can be placed in the target rack at the bottom to the right of the position rack. Pushing the target block will append the selected media element to its contents, which is confirmed visually on the display and also by acoustic feedback and a haptic "click". Pressing and holding the target block, while at the same time moving the source block to another position will append a range of elements, in analogy to dragging a selection with a mouse. Similarly, the sequence rack allows to combine media contents into a single sequence, which can be bound to a target block placed on the target rack.

3.2.1.4 The Tuister

The Tuister is a TUI for navigating hierarchies, such as nested menus. It is designed to support bimanual interaction and provides input and output capabilities at a single device [Butz et al., 2004].

The Tuister device consists of two cylindrical parts on a common central axis (see Figure 3.12). The right cylinder is called the head and contains several displays, for example six displays in a hexagonal arrangement as shown in Figure 3.12. The left part is called the handle and can be used for manipulation.

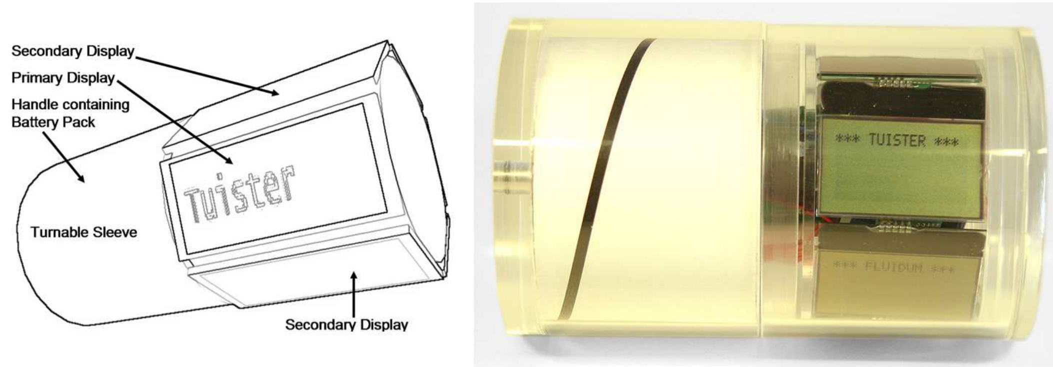


Figure 3.12*: The Tuister design (left) and prototype (right) (sources: [Butz et al., 2004])*

The user can turn the head of the Tuister to choose from a list of items that is shown on the displays. Displays on the backside can be switched off and reappear at the front with different content, such that it can be scrolled through arbitrarily long lists. To increase scrolling speed, the user can give a spin to the head, which will rotate freely until the spin fades or the user decides to stop the movement when the list is close to the item that is being searched for. Once the target appears at the primary display that is directly facing the user, it can be selected by rotating the handle clockwise while the head is held fixed by the other hand. If a submenu is selected in this way, it will be entered, while rotating counterclockwise will move one step up in the menu hierarchy. Letting the handle spin freely would go down along a set of default choices and spinning it counterclockwise up to the root. The creators of the Tuister envisioned the device to serve as a personal multi purpose device for instrumented environments, which the user can bring along as a universal interface for arbitrary applications that require navigation through hierarchical menus.

3.2.1.5 The I/O Brush

The I/O Brush is a device that looks like a regular paintbrush, but it is instrumented with a video camera, touch sensors and LEDs [Ryokai et al., 2004]. These sensors allow users to pick up arbitrary colours, textures and movements on surfaces from anywhere in the environment by touching an object with the head of the brush, in analogy to dipping a paintbrush into a colour pot. This virtual ink can now be drawn onto a special canvas, a touch sensitive back projection display, as seen in Figure 3.13.

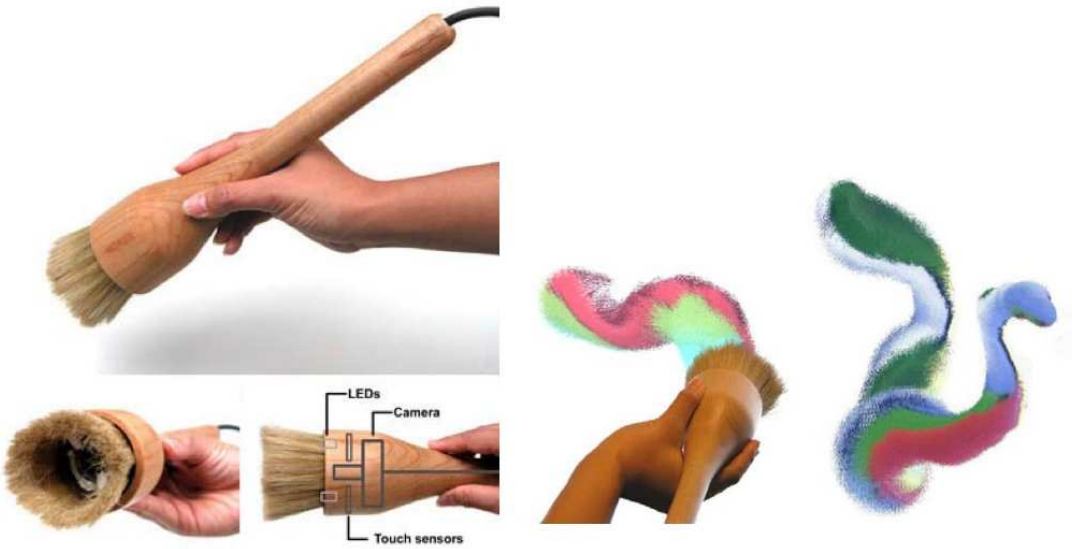


***Figure 3.13**: The I/O Brush (left) in action (right) (source: [Ryokai et al., 2004])*

The I/O brush was designed as a tool for children to explore colours, patterns, and movements found in their environment by sampling and applying them to virtual drawings. Force sensors embedded at the bottom of the brush bristles detected pressure applied on the brush. Thus, when in sampling mode, the LEDs inside the hollow bristle would illuminate the target to enable the camera to sample the surface. In drawing mode, the system would paint the sampled image or motion (done by sampling over a longer duration) when pressure on the bristle is detected. Artists can control the size of the sampled image drawn on the canvas by the amount of pressure applied on the brush.

Using the I/O brush closely resembles the actions performed in a real painting situation, realising a very intuitive and effective interaction. In addition to that, the sampling feature provides a function, which is otherwise difficult or impossible to perform for novice artists with traditional painting utilities. Although the brush is theoretically still usable as a common brush, it is not designed to retain its original, analogue function, as this would interfere with the integrated technology and also require a conventional canvas.

3.2.1.6 TUIs for Media Control

In 2005 we have tutored a TUI design course, in which a collection of TUIs for media control in a home scenario [Butz et al., 2005] was created. This interdisciplinary student project involving product designers and computer scientists showcases a diversity of interaction metaphors suitable for the rather simple and well defined task of controlling the main functions of a HiFi system at home. The development process included 2D sketches, the construction of low-fidelity hardware prototypes, from which the most appropriate solution was selected and prototypically implemented. All implementations were either based on the Smart-Its sensor platform [Beigl and Gellersen, 2003] [Holmquist et al., 2004a] or realised with custom made electronic circuitries.

The Hanging Twines constitute a chandelier-like installation, which would hang from the ceiling and would be part of the environment, similar to a piece of furniture or a lamp. Plexiglass rods were chosen as the material for the twines, which are illuminated from the top end and thereby achieve luminosity over the whole length. The ends of the rods were designed to provide affordances or at least visual invitations for pulling and/or turning (see Figure 3.14), which would activate its function.

***Figure 3.14**: The Hanging Twines*

The rods are shaped in such a way, that they would reveal their functionality by a similarity of their shape to the conventional symbols for media control, such as play, pause, forward, rewind and stop. Further, symmetric functionality such as rewind and fast forward could be mapped to one rod, since the symbols only differ in their direction. In consequence, the current orientation of the rod would further define which of the two actions will be executed when the rod is pulled.

Another solution is called Pyramid Control, a sphere consisting of four ”wings” and residing on a stand with a LED, which indicates the front side of the device [Bartels, 2004]. The LED should face the user, such that the spatial mapping of the wings are unambiguous. In its initial closed position the dividers of the movable parts indicate the partitioning and, in conjunction with a small cavity on top of each wing, suggest the possible interaction. This kind of foldable device allows direct manipulation of player functions by adjusting the movable parts - giving an additional visual feedback of the system state by the current wing positions, see also Figure 3.15.

Figure 3.15*: Form studies (on the left) and the final model of the Pyramid Controls*

The prototype utilises the acceleration sensors of Smart-It particles, which were integrated into each wing, in order to detect the positions of the wings and to activate functions of the player accordingly. A closed object means stop, pulling the front side down starts playing, while half-way down pauses the playback. The backside directly controls the volume by its position whereby closing this side means ”mute”. The left and right wings rewind and forward respectively when pulled down half-way while by tilting them down further tracks are skipped in the corresponding direction.

A similar approach is taken by the Flower, a stationary device with three layers of switches circularly arranged around the body of the product. Two switches were connected to each of these petals to detect up- and downward movements, returning a soft click as feedback. Each button sends a signal to a controller chip when it is pressed and the controller chip encodes the state into an infrared signal and activates the diode to transmit it until the button is released. The outer layer control overall volume or sound processing, the middle layer could switch between components of a HiFi system, while the inner part controls playback on a single component (see Figure 3.16).

The last tangible interface for media control created by this project is Flip’n Twist, a bimanual device, consisting of two different parts, which could be used separately [Taffin, 2004] (see Figure 3.17). The cube is used to set the system into the desired operation mode. The different modes can be selected by flipping the cube from side to side, where the selected mode faces up. One typical example is to flip the cube until the upper

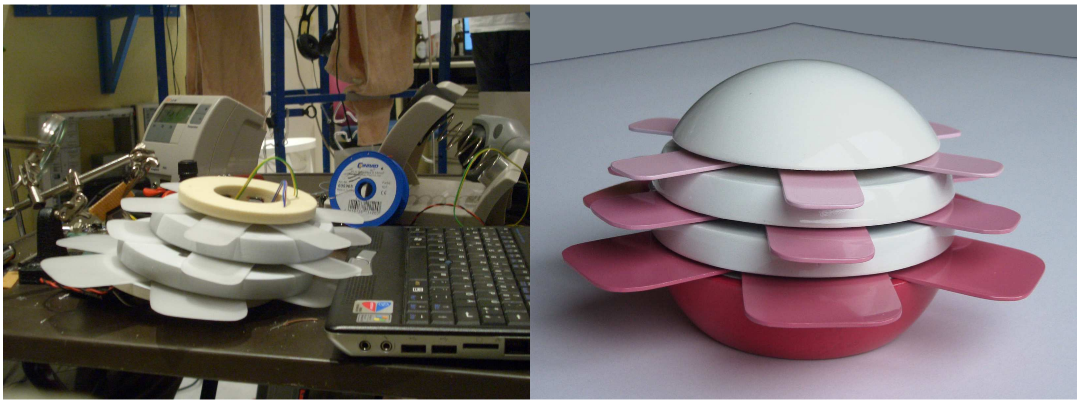


Figure 3.16: *Work in progress (left) and the final prototype of the Flower*

face displays the symbol for volume adjustment. The exact sound level is then selected by twisting the second device.

Although the twisting device was designed explicitly to input continuous data, such as volume, it can also be used to make choices or a binary decision by left or right turns. This helps for example to implement functionalities provided by traditional remote controls for HiFi equipment, e.g. skipping or repetition of the current song. While the cube is restricted to be used on a planar surface for correct orientation detection, the twisting device can be used in any position, e.g., comfortably from the sofa. The combination of the cube and the twisting device allowed seven discrete selections (skip song, repeat song, play, pause, stop and shuffle) and four continuous selections (speed of fast forward, speed of rewind, increase volume, decrease volume).

3.2.1.7 The reacTable

The reacTable combines tangible interaction with multitouch interaction on a tabletop interface, realising a musical instrument designed for installations and casual users as well as for professionals in concerts [Jordà et al., 2005] [Jordà et al., 2007]. It aimed at providing immediate and intuitive access and at the same time the flexibility of rich digital sound design algorithms. One or more musicians can share the control of the instrument by rotating or moving physical artefacts on the surface, constructing different audio structures as in a tangible flow controlled synthesiser.

The main sensory component consists of an infrared camera below the table surface, which is in charge of the recognition of fingertips and artefacts that have so called fiducial markers attached underneath. Fiducial markers are optical markers, i.e. graphical symbols that can easily be identified and tracked in real-time in a video stream.

Each reacTable item represents a synthesiser component with a dedicated function for either the generation, modification or control of sound. Following a set of rules these objects

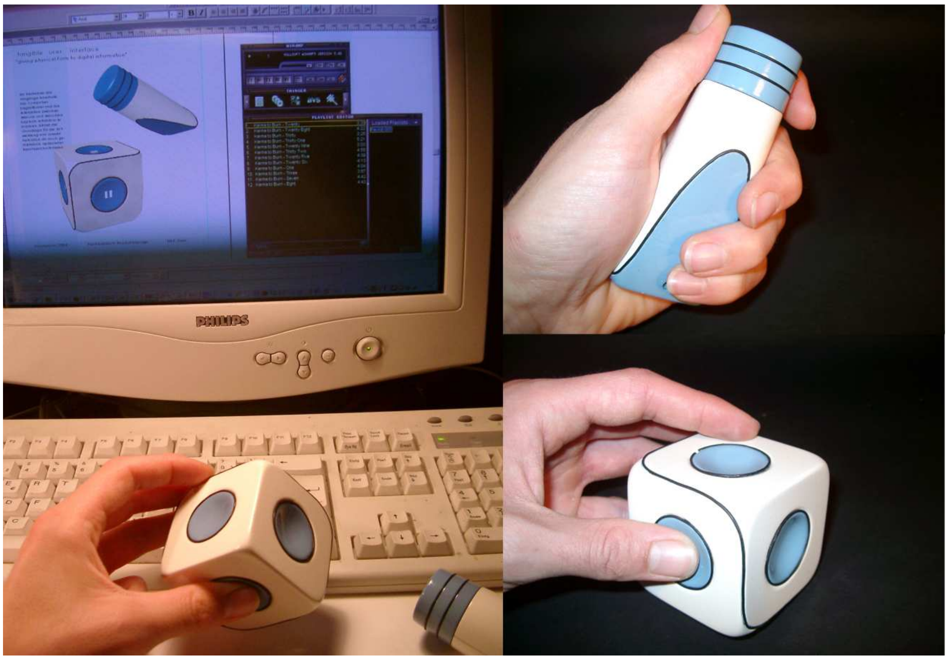


***Figure 3.17**: The Flip'n Twist prototype, a bimanual device*

are automatically connected and disconnected, according to their type, affinity and proximity with neighbour objects. The resulting topology is permanently visually represented on the table surface as shown in Figure 3.18. Circles around the physical objects carry information about their behaviour, parameters values and configuration states, while the connecting lines represent the sound flow produced or modified at each node.

Therefore all shapes, lines or animations drawn by the graphical component convey certain states or attributes: Control lines indicate the intensity of the values they transport; vibrations of metronomes and similar objects are animated like heartbeats. The white 180 degree circular fuel gauge surrounding an object indicates its rotational values and the dots show the position of the second parameter slider (see Figure 3.18). Fingers can also be used to temporarily mute audio connections. Muted connections, which are represented with straight dotted lines, are reactivated by touching again the previously muted object.

3.2.2 TUIs in Multimodal Interfaces

During the next section we will introduce tangible interfaces that involve multiple modalities, which also includes multimodal user interfaces incorporating the haptic input modality. Compared to traditional TUIs as presented in the previous section, the foci here are shifted from a purely tangible interaction style towards the combination with other in- and output modalities.

***Figure 3.18**: The reacTable (source: [Jordà et al., 2007])*

3.2.2.1 The Rasa System

The Rasa system was implemented for military command posts in field training exercises and enables officers to use to maintain their common practice using paper maps and Post-it notes to support military command and control tasks [McGee et al., 2000] [Cohen and McGee, 2004]. During simulated battles, the officers track locations and activities of friend, foe and neutral parties on a paper map by drawing unit symbols on Post-it notes and positioning them on the map. The unit symbol is derived from a composable language for these pictograms that is learned during officer training and used daily. It depicts the unit's strength, composition, and other relevant features. When new positions are reported over radio, the notes are placed to the new locations of the units, such that the the map provides an accurate overview of the current state of the battle, enabling superiors to make critical decisions quickly.

The goal of Rasa was to augment the tools in an existing work practice and preserving their original use as much as possible. The resulting system tracks the main user actions, the creation and manipulation of Post-its and their movement on the map, and also offers supplementary functions such as extending the information about units represented by Post-it notes through voice annotations or gestural input. The system reflects the complete battle state and is therefore able to distribute information between the involved command centres. The information flow of Rasa is illustrated in Figure 3.19.

With Rasa the battle map is registered to a large touch-sensitive tablet, and the Post-its

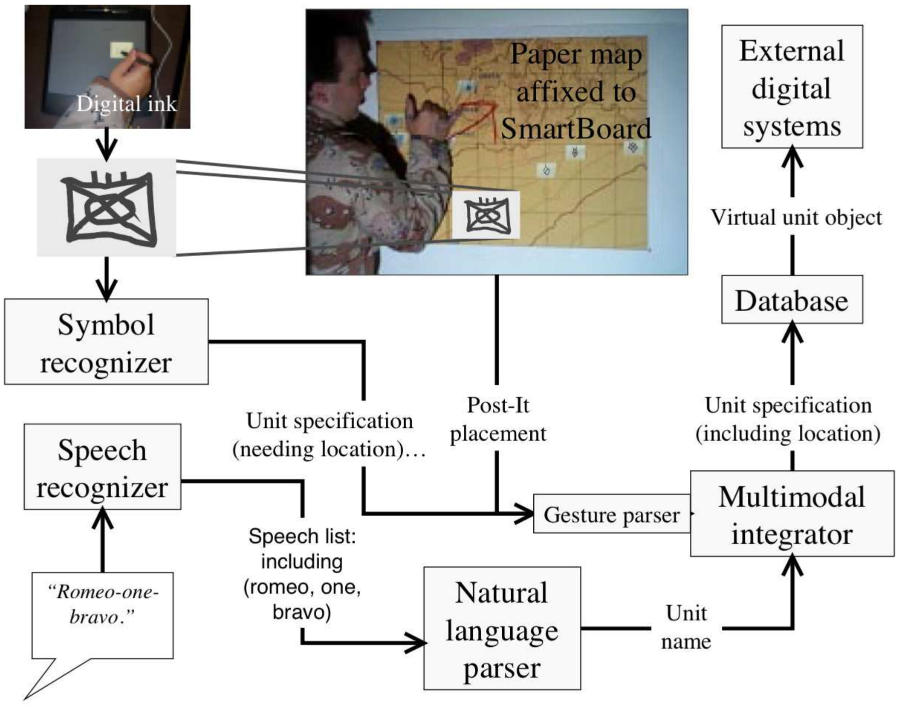


Figure 3.19: *The Rasa system architecture (source: [McGee et al., 2000])*

initially rest upon a tablet that recognises which military unit symbol is drawn on a Post-it note. While drawing the symbol the user can also speak identifying information about that unit. The computer recognises both the symbol and the utterance, fusing their meanings using multimodal integration techniques. When the user places the note onto the paper map, the unit and its data such as the position is recorded in the system's database. If a report arrives over radio indicating that the unit has moved, the user still only needs to pick up the note and move it to the new location on the map. The user can further augment the note on the map with speech, gesture, or both. For instance, drawing an arrow starting at the centre of the note while saying that the unit is moving in this direction at 20 kilometres per hour will trigger Rasa to project this new information onto the paper map as an arrow labeled with ”20 kph”.

3.2.2.2 The Mobile ShopAssist

Another important example for multimodal interaction integrating real items in instrumented environments is the MobileShopAssist (MSA) [Wasinger and Wahlster, 2005] [Wasinger et al., 2005]. The MSA is a personal shopping assistant running on a mobile device and utilising RFID-based infrastructure of the (shopping) environment. The functionality of the prototype comprises the query of attributes of products and the comparison between two product items. The research focus of the MSA project was on rich symmetric multimodal interaction, thus the shopping assistant prototype provides the input modalities speech, handwriting, gesture, and combinations (see Figure 3.20). Two kinds of gesture inputs are feasible: Intra-Gesture refers to the interaction with virtual products and occurs when a user touches a virtual product on the touch-screen. Extra-Gesture refers to the selection of real products equipped with RFID tags, by picking them up from the shelf.

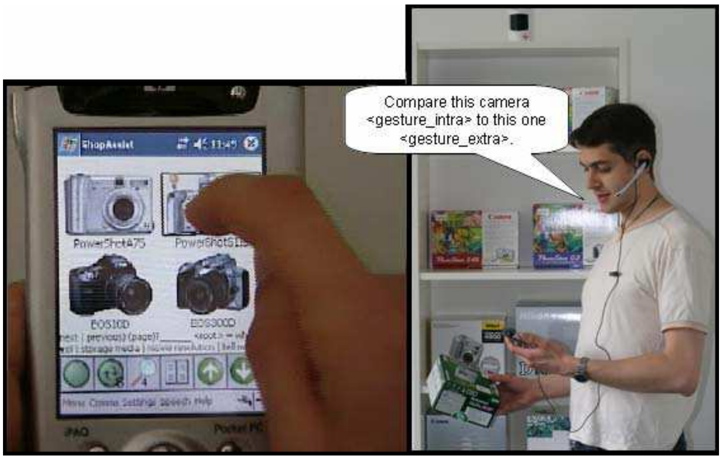


***Figure 3.20**: The MobileShopAssist*

Thus, a user interested in buying a digital camera would for instance approach a shelf, which will automatically synchronise its contents with user's mobile device. After synchronisation, the user may ask about the product's attributes, e.g. by saying "What is the optical zoom of this camera?" and taking the product out of the shelf, which she or he is interested in. Alternatively, the exactly same query could be triggered by clicking on the symbol "optical zoom" and uttering the type and model of the respective camera0. A user study conducted in an electronics store has revealed that the most preferred choice from all possible modality combinations is the combination of speech and extra-gesture, probably because this interaction procedure resembles the interaction with a real sales person. The authors also pioneered in applying a spoken dialogue interface to inanimate objects, which means that instead of asking "What is the price of this camera?" a user would say "What is your price?" towards a digital camera. The user study has revealed that users dislike to directly converse with a bar of soap but would be generally willing to talk to digital cameras, cars or personal computers [Wasinger and Wahlster, 2005].

3.2.2.3 The Augmented Dorfladen

Another project in the shopping domain is a multimodal product information and comparison system that we have designed for a rural corner store (the so called *Dorfladen*) [Schmitz et al., 2009] [Schmitz and Quraischy, 2009] [Quraischy, 2009] [Minina et al., 2009]. Such a Dorfladen only provides a limited assortment, which is extended by the system with virtual items on embedded displays, allowing customers to inspect and select unavailable products along their real world shopping activities for later ordering (see left part of Figure 3.21). The user interface combines tangible interaction and natural language dialogues to interact with real and virtual products in one instrumented space.

Two main functions comprise the interactive functionality of the prototype, product inspection and comparison: The product inspection function supports customers to inspect single products by displaying basic product information such as the price and nutrition facts in a consistent fashion. To request information about a real product the user merely has to pick up the product from the shelf. Afterwards all of the product information is listed on the screen (see Figure 3.21 on the right) as a visual response. Large letters help elderly customers who have difficulties in reading the original food packages. Whereas information about real products is often redundant, since it is printed on the product itself, virtual products are not present and require the use of this function to acquire additional information. Selecting a virtual product happens through touching its picture on the screen, which results in the same response as if a real product were selected.

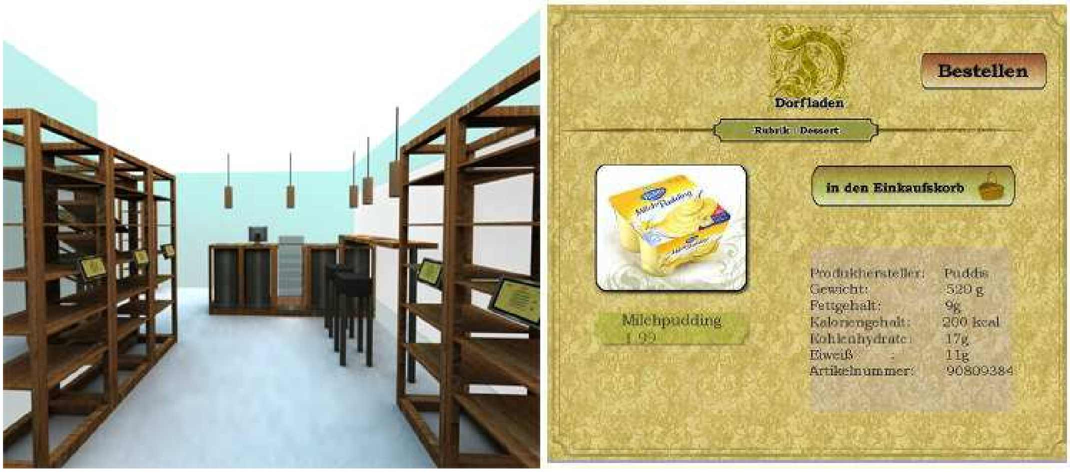


***Figure 3.21**: The Augmented Dorfladen: Interior design layout (left), product inspection page (right)*

In order to accelerate the process of finding the right product the comparison of more than just two products is possible as well. In this comparison, real and virtual products can be arbitrarily mixed. The prototype is built to handle two kinds of comparisons: With tangible selection users can select virtual items by touching their pictures on the screen and real items by taking them out of the shelf. Tangible selection can further be complemented by spoken specification: If a user, for example, wants to find out, which products contains the least amount of fat, speech can be used to specify this request. The visual response will be complemented (bold font and green marker) to indicate the product that complies to the request. Furthermore speech output is generated.

The following example interaction illustrates how four products are compared with speech input: The User touches three virtual products one after another and takes out one real product, at the same time she asks ”Which product has the least amount of fat?”. The system displays the response as a comparison, listing all selected products with their attributes. The answer is marked with the green point and the system renders the utterance: ”The product xyz has the least amount of fat, it contains 1 gram.”

3.2.3 TUIs Combined With Life-Like Characters

This section covers the few instances of interaction research that attempt to combine concepts of TUIs with that of life-life characters, and thus apply both tangible and anthropomorphic interaction principles.

3.2.3.1 Virtual Constructor aka COHIBIT

Figure 3.22: *The COHIBIT installation*

The COHIBIT system (**CO**nversational **H**elpers in an **I**mmersive exhi**BI**t with a **T**angible interface) was developed as an edutainment installation for theme parks, providing a robust and simple interface while at the same time fostering creativity and fun [Ndiaye et al., 2005]. The user is confronted with two life-sized characters on a large screen behind a set of instrumented 3D puzzle pieces, which resemble car parts and invite to play with, i.e. to assemble a complete car. The main idea is that assembling actions of the user are tracked and commented by the two life-like characters, such that visitors get the feeling that the characters observe and understand their actions and provide feedback or guidance (see Figure 3.22). During the construction phase users can pick up pieces from the shelves and put them on the assembly bench. They can further rearrange the order of pieces or replace them with other pieces. The virtual characters play different roles, e.g. as guides by giving context-sensitive hints about how to complete the current construction.

Several RFID tags are invisibly embedded into each puzzle piece such that position and orientation of an item on the assembly board can be inferred. This approach simplifies the interaction tracking to the detection of a limited amount of discrete states and ensures the robustness required for such an installation. It also allows multiple users to interact with the system and move car pieces. By using physical objects for the car assembly task, users can therefore influence the behaviour of the two virtual characters without being aware of it. In summary, the COHIBIT installation provides a tangible input channel through the physical car pieces and responds in a multimodal manner via coordinated speech, gestures and body language of the virtual characters.

3.2.3.2 IDEAS4Games: An Instrumented Poker Installation

A similar approach as in COHIBIT was chosen by Gebhard et al. [Gebhard et al., 2008]: Their system uses real poker cards each equipped with unique RFID tags and allows to play draw poker against two virtual characters - one cartoon-like, friendly looking character and a mean, robotic counterpart. The user acts as the card dealer and also participates as a regular player (see Figure 3.23).

The poker table has three outlined areas for poker cards: one for the user and one for each virtual character. RFID sensor hardware is placed under each area, such that the system detects where the cards a placed in order to assign them to the corresponding player. The virtual characters as well as other information relevant for the game are displayed on a large screen behind the table, as shown in Figure 3.23.

When a user initiates a game, the characters explain the game setup and rules and direct the user to shuffle and deal the cards. During the game the two virtual characters react to events triggered by RFID-tagged cards or timeouts, e.g. when the user deals the cards too slowly. The overall system aims at intuitive interaction through the use of real poker cards and thus preserving familiar game activities, combined with consistent and convincing character behaviour, realised with real-time affect computation based on game events and expressed through the character's speech and body movements.

3.2.4 Summary

We have seen different types of tangible interfaces in this section, beginning with typical TUIs that concentrate on physical manipulation of interface artefacts, followed by multi-

***Figure 3.23**: The A.I. Poker installation at the CeBIT fair in 2008 (source: [Gebhard et al., 2008])*

modal systems and tangible interfaces combined with life-like characters.

Classical TUIs typically employ application specific physical interaction artefacts that lose their semantics outside the application they were designed for. Thus, by definition, such artefacts do not belong to the category of smart objects anymore. A border case is the I/O brush, which retains visual and also most other physical properties of a real paintbrush, but loses its original functionality due to the integration of sensory technology. But this prototype reveals the apparent advantage of exploiting the (learned) affordances of known physical tools, which, in this case, enables both children and adults to intuitively apply the brush-like interaction style in this drawing application. Other TUI instances, such as the Hanging Twines, show how physical arrangement and shape can implement appropriate mappings and affordances.

The multimodal systems introduced in Section 3.2.2 attempt to realise natural human computer interaction by building on distinct cognitive human skills of processing and expressing multiple in- and output channels in parallel. Speech is often a the primary information channel, but tangible interaction always plays a major role in such systems as well. The combination of life-like characters with tangible interaction in entertainment and exhibition scenarios underlines the potential of inspiring and entertaining users with such an approach, although the anthropomorphic and tangible elements of the introduced user interfaces are still completely disconnected.

3.3 Anthropomorphic and Animalistic Interfaces

The following section introduces related projects that involve different types of anthropomorphic user interfaces. We start with embodied conversational agents, the most prevalent type of applied anthropomorphism, before we shift towards anthropomorphism in physically embodied interfaces, such as smart objects and robots.

3.3.1 Embodied Conversational Agents

Embodied conversational agents (ECAs) are a type of intelligent user interface that are typically represented graphically by human or animal bodies in a life-like and believable manner. Besides recognising and generating verbal and nonverbal output, ECAs apply speech, gestures or facial expressions to mimic typical functions and properties of human face-to-face conversations, such as turn-taking or feedback [Cassell et al., 1999].

3.3.1.1 Virtual Human

The Virtual Human project is a joint research project conducted by 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. The overall goal of the project is to facilitate realistic anthropomorphic interaction agents by combining state-of-the-art computer graphics technology with speech and dialogue processing [Göbel et al., 2004]. The dialogue and narration engine acts as a control unit of the Virtual Human run-time environment, while computer graphics technologies were utilised for photo-realistic rendering of the Virtual Human characters. Instead of using pre-scripted dialogs Virtual Human agents are driven by a behaviour engine, which is in charge of coordinating speaking turns dynamically.

The modular architecture of the Virtual Human platform provides a high configurability of the system, allowing for an efficient setup of new application prototypes, depending on the domain and application tasks (see Figure 3.24).

The authoring environment enables developers to configure application scenarios by defining stories, agents and their behaviour towards each other and the user. The main purpose of the narration module is the selection and concatenation of scenes at run-time based on declarative story models. Thus, the module controls the narrative progress by sequentially choosing a scene out of the pool of available scenes, which are the elementary building blocks of the story line. The dialogue module creates scripts which contain detailed action and dialogue descriptions such as timing for emotions, lip-synchronisation, gestures or facial

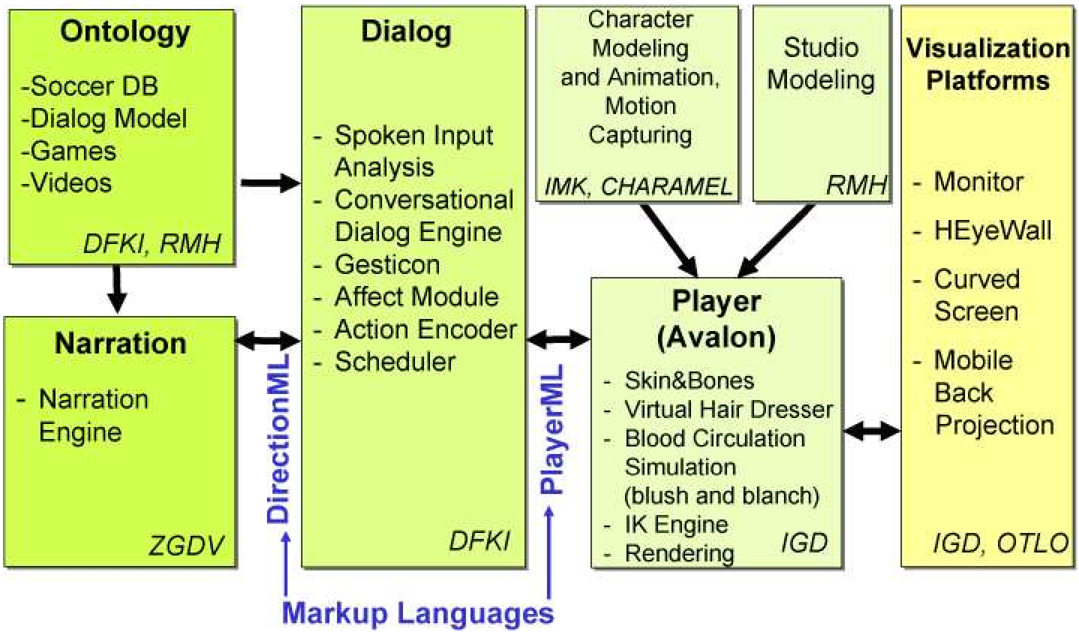


***Figure 3.24**: The Virtual Human architecture*

expressions. The resulting script is interpreted by the player module (an extended version of the Avalon player), a rendering engine that handles the scene graph by defining geometric, graphical and behavioural properties.

One Virtual Human scenario comprises a soccer game show for two users . One moderator and two soccer experts are rendered as virtual agents and interact with the users, who are standing in front of the life-size projection behind a microphone/trackball setup. The moderator presents video clips of football scenes, which are stopped before the result of a scene is revealed, for instance when a forward attempts to shoot a goal. The contestants can now make a guess about what might happen next or ask the experts for some advice (see Figure 3.25). This procedure is repeated three times, before the moderator presents a final evaluation and the winner of the this first part of the game.

The winner can now proceed to the second phase, where she or he is requested to put up a lineup for the German national team with the help of the moderator and one of the experts. An exemplary dialogue (translated from German) might happen as follows:

(1) Moderator: Ok, let's get started.
(2) User: Put [characters gaze at user] Oliver Kahn up as keeper.
(3) Expert Herzog: [nods] That's an excellent move!
(4) Moderator: [nods] Great, Kahn as keeper.
(5) User: Miss [characters gaze at user] Herzog, give me a hint!
(6) Expert Herzog: [smiles] I would definitely put Ballack into the central midfield.

***Figure 3.25**: The Virtual Human demonstrator at the CeBIT fair in 2006)*

(7) User: Ok, [characters gaze at user] let's do that.
(8) Expert Herzog: [smiles] (nods) You won't regret this move.
(9) Moderator: (nods) Great, Ballack as central midfielder.
(10) User: ... [hesitates]
(11) Moderator: [encouraging gesture] Don't be shy!
(12) User: Hhm, [characters gaze at user] put Metzelder to Ballack's left. [. . .]

Interesting to note in this example are the utterances, which require information about discourse and context, e.g. in line 12, when the user performs a spatial reference to an element of the scene. This requires the system to be able to determine which location is denoted by "Ballack's left". The gazing behaviour also requires an awareness of each agent about the current role of all participants, including the user.

3.3.1.2 The Virtual Room Inhabitant

The Virtual Room Inhabitant (VRI) is a virtual character that is capable of freely moving along the walls of an instrumented environment and is designed to facilitate intuitive interaction with systems located in this environment [Kruppa, 2005] [Kruppa et al., 2005]. The character is aware of the users position and orientation within the room and is able to offer situated assistance as well as unambiguous references to physical objects by means of combined gestures, speech, and physical locomotion. The VRI is realised through a steerable projector and a spatial audio system, which position the character within the environment visually and acoustically. Figure 3.26 shows the character on a wall besides a large touch screen.

One novel approach in this work is the transfer of the concept *deictic believability* from virtual 3D worlds to an augmented physical world, by allowing a virtual character to move

***Figure 3.26**: The Virtual Room Inhabitant*

within physical space. Through a steerable display the character is able to appear onto any surface within the room, such that it can refer to physical devices through gesture and speech. For example, the VRI can reside on a wall close to a bookshelf and utter ”To my left you can see the bookshelf, which...” and at the same time point to the correct position of the shelf.

The main idea of the example application is to allow the character to appear as the host of the environment, that is always aware of the state of each device and service, such that it can provide situated assistance. The VRI is capable of welcoming a first time visitor and its main purpose is to explain the setup of the environment and to help users while interacting with the instrumented environment.

3.3.1.3 FitTrack - A Relational Agent

The main purpose of the FitTrack project [Bickmore et al., 2005] [Bickmore and Picard, 2005] was to examine the concept of *relational agents*, which are *”computational artefacts designed to establish and maintain long-term social-emotional relationships with their users.*” Thus, the goal of the FitTrack relation agent was to construct, maintain, and evaluate such a relationships between the user and the agent.

The MIT FitTrack system was designed to be used on personal home computers on a daily basis during a one-month study, with each interaction with the system lasting approximately ten minutes. Laura, the relational agent played the role of an exercise advisor, which the participants can talk to about their physical activity. The system design was mainly based on studies of interactions between professional exercise trainers and their clients, surveys of representative study participants, literature reviews of therapist-client and physician-client interactions.

To be able to support a wide range of personal computers, a client-server architecture was developed in which the client was running on each participants' computer and realised as lightweight as possible. It consists of two web browsers coupled with a vector-graphics-based embodied conversational agent and synchronised with a text-to-speech engine. All dialogue and application logic was kept on the server.

Although the agent was able to render synthesised speech and synchronised nonverbal behaviour, users primarily contributed to a dialogue by selecting text phrases from multiple-choice menus, which were dynamically updated depending on the conversational context (see Figure 3.27). One advantage of this design decision was that by constraining what the user can say in each state of the dialogue, the responses of the agent can be constructed to address all possible inputs in a meaningful manner.

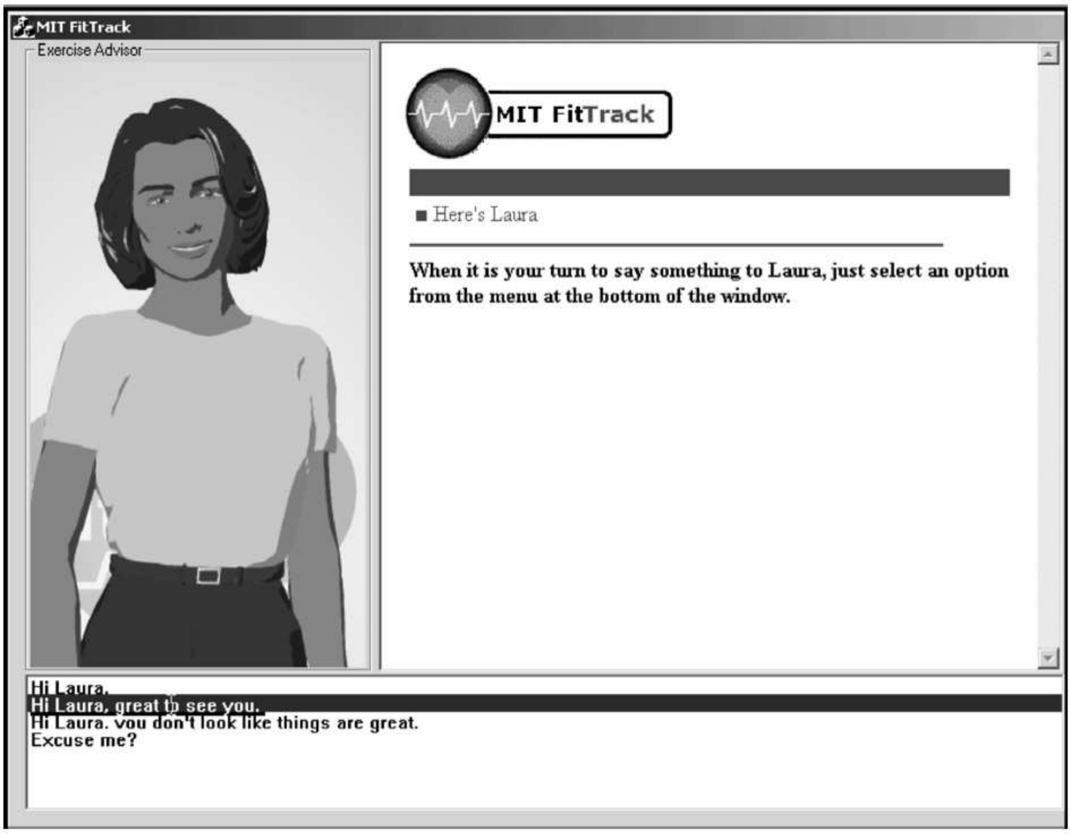


***Figure 3.27**: The FitTrack interface (source: [Bickmore et al., 2005])*

The embodied agent was able to perform range of nonverbal behaviours for communicative and interactional functions, including hand gestures, body posture shifts, gazing at and away from the user, head nods, and walking on and off the screen. It also supported four different facial expressions, variable proximity of the agent to the camera and several idle-time behaviours. The applied text-to-embodied- speech system supported the exercise dialogues, by, for example, allowing to specify the conversational frame (either task-oriented, social, empathetic, or encouraging) in the script, which would be automatically manage appropriate changes to facial expression, proximity and speech synthesiser intonation.

The relationship model applied by the FitTrack agent is basically a stage model, in which the change in relationship is fixed over the 30 scripted dialogues. Since the working alliance is thought to increase to a stable level after approximately seven sessions in psychotherapy, the relational strategies of the FitTrack agent take place over the first seven interactions accordingly. The relational strategies applied in the interaction dialogues include social dialogue, empathy dialogue, meta-relational communication, humour, continuity behaviours, and appropriate forms of address and politeness. Nonverbal behaviour was also modulated to reflect high or low immediacy, adjusting the frequency of hand gestures, eyebrow raises, head nods, and gaze-aways, as well as the proximity of the agent (e.g., the agent would appear to move closer to the user in a high immediacy condition).

To evaluate the relational qualities of the FitTrack system, it was compared to an alternative version of the agent that did not employ any relational strategies. Both versions were exposed to the study participants over 30 days, which were surveyed in questionnaires afterwards. To sum up the results, it could be shown that subjects in the relational condition reported that they liked the agent significantly more than those in the non-relational group, while at the same time a closer relationship with Laura was reported as well by the participants in the relational condition. Furthermore, this group responded much more in favour of continuing working with Laura than the non-relational group.

3.3.2 Kinetic Smart Object Interfaces

In this section we will introduce a fairly novel category of interfaces, which explores kinetic motion to express life-like behaviour in smart objects.

3.3.2.1 Living Interfaces

The Living Interfaces project of the Deutsche Telekom Laboratories explores in series of experiments and prototypes how reduced life-like movements can be beneficial for interaction with everyday objects. The idea of the Impatient Toaster [Burneleit et al., 2009] is to create sympathy and a closer relationship to a kitchen appliance by enabling the device to indicate needs by certain movement patterns. This means in particular that the toaster starts to shake nervously after a longer period of inactivity in order to remind the dweller of eating something. Once bread is inserted into the toaster and the start button is pressed, it calms down. Singular, subtle movements during the toasting period signalise activity. After the toasting phase the bred moves upwards such that it can be taken out, while the machine again moves excitedly until the toast bred is removed. The prototype toaster was not fully implemented, instead it was equipped with servo motors, which have to be manually started, such that an informal user study could be conducted in a Wizard-of-Oz setup. Six persons were exposed to the impatient toaster without being told that the device was already active, instead they were instructed to fill in a questionnaire when the toaster suddenly started to shake. The observed reactions were similar: After they were startled at first, the participants approached and touched the toaster and began talking to it, asking what it needs or wants, as shown in Figure 3.28. When the toaster wiggled again, after it has been already calmed down by feeding it with toast bread, all subjects returned and started talking to and touching it attempting to calm it down. Although this study does certainly not comply with formal empirical methods, it indicates the potential of inducing joy and an anthropomorphic stance towards an object with rather simple means.

Figure 3.28*: Study participants touching the Impatient Toaster to calm it down (source: [Burneleit et al., 2009])*

Similarly, the Thrifty Faucet [Togler et al., 2009] project sought to explore with a pilot study how movement can induce a life-like impression on users and whether this can be utilised to invoke an emotional reaction. The proposed use case scenario here was to create awareness, e.g. for water consumption. The Thrifty Faucet is a bendable plastic tube guided by three pairs of steel wires driven by servo motors, such that the body and head of the pipe can move independently and take up several distinct poses, as shown in Figure 3.29). For the study 15 live motion patterns were imagined from interaction possibilities with a water-tap, covering expressions from positive feedback to denial. These patterns were demonstrated to 9 participants, who were handed out questionnaires afterwards asking about their impression on the presented interface. The overall reaction inclined to be positive and intrigued, while the reported emotions ranged from fright to amusement. Making the pipe ”look” / ”point” at the user achieved particular high attention, amplified by the fear of being splashed with water, as many subjects reported.

The idea of the The Pulsating Mobile Phone is to exploit the human ability to perceive

Figure 3.29: *The Thrifty Faucet in different postures: seeking, curious, rejecting. (source: [Togler et al., 2009])*

and interpret signs of human life [Hemmert and Joost, 2008]. In this case a vibration pattern supposed to mimic a pulse is applied to a mobile phone for delivering information about the presence of missed calls or unread messages. Whenever the phone is in normal status (i.e. no missed events, good network reception, battery is not going to be empty soon etc.) it shortly vibrates twice every few seconds, counterfeiting a calm heartbeat (see also Figure 3.30). Two versions of the phone's reaction to certain events such as a missed call were examined: The first idea was to change the phone's heartbeat pattern into an excited mode, causing the phone to vibrate more frequently. A second instantiation of the prototype replaced the excitement with absolute silence. The hypothesis of this work was that the calm pulse would be habituated after a certain time, not distracting the user and providing for a subtle awareness of the phone's state. It was further expected that a sudden change or absence of the vibration would be immediately recognised.

Figure 3.30: *Different prototypes of the Pulsating Phone (source: [Hemmert, 2008])*

The results of first explorative studies were mixed for both alternatives: While some users got used to the pulse vibration, most found it increasingly annoying and distracting - depending on activity and context. Obviously a potentially obtrusive and continuous tactile display has to be implemented carefully. A further interesting finding was that participants of the study reported some kind of *"psychological gap left by the phone when taken out of the pocket"*, such that the vibration could not be perceived anymore.

3.3.2.2 Kinetic, Animalistic Products from Art and Design

Animalistic products also became of interest for artists and designers, reflecting on a variety of topics which might or might not be related to autonomic objects. Work in this field does often not intend to achieve real world applicability and is certainly detached from the need of adhering to scientifically valid methods. But as such projects may also be a source of inspiration, we include a few representative installations into this state-of-the-art overview.

The Shy Surveillance Camera [3] - or also called (In)Security Camera - is a security camera that is able to recognise and follow people in real-time through a pan and tilt unit on the hardware side in combination with computer vision software (see also Figure 3.31, left). Relating to the significant increase in the deployment of electronic surveillance systems in public space, the idea of this project is to reflect on the diminishing esteem for privacy and in fact to invert this situation. Thus, the ”behaviour” assigned to this camera appears to be of that of an insecure personality. It is easily startled by movements, shy in the presence of strangers and tends to avoid direct eye contact. This is expressed e.g. by sudden reactions to detected movements, aiming at the centre of a moving object and immediately turning away, if the face of a person is recognised. Interesting here is to see how convincingly a human-like attitude could be achieved with very simple means: Showing attraction to movement by letting the camera point to it and illustrating shyness by turning the camera away from human faces.

***Figure 3.31**: The (In)Security Camera (left), the Luxalive Lamp (right)*

The product designer and creator of the Luxalive lamp[4] (see Figure 3.31) aims at socially empowering people by assigning different characteristic traits to a lamp, which control its behaviour when in use. The lamp is attached to a robotic arm and acts in two modes: If the user is supposed to be dominant, the lamp behaves cautious and polite by optimising its brightness and distance to the book being read in order to let the user feel in control. The other alternative is tailored for selfless people that enjoy helping others. In this case the lamp ceases to work and slumps down until the user tilts it up and rubs the head, supposedly to convey the feeling of being needed. The basic idea of choosing different personalities for household appliances

[3] `http://www.vitagrrl.com/art`, last visited October 20, 2010
[4] `http://www.coroflot.com/zoontjens/Portfolio1/2`, last checked October 20, 2010

to match the owner's personality by creating a social fabric between user and object with distinct dependencies is generally appealing and provides an alternative approach of choosing virtual personalities to the concept of *similarity attraction* [Byrne and Griffitt, 1969], which basically states that the degree of similarity correlates with the attraction towards each other.

3.3.3 Shape-Shifting Smart Objects

In this section we will introduce work on smart objects that are able to change their shape or other aspects of their outer appearance and thereby potentially create the impression of an organic or living entity. As a matter of fact, in one of the examples that will be presented the interface medium *is* organic.

3.3.3.1 The Inflatable Mouse

The prototype of the inflatable mouse [Kim et al., 2008] employs a common off-the-shelf computer mouse that is instrumented with sensors and actuators around a small balloon attached on top of the actual mouse (see Figure 3.32). Several touch sensitive sensors are placed around the balloon and also at certain spots where traditional mouse interaction takes place, i.e. for left and right mouse clicks and the scroll wheel in between. The balloon itself contains a pressure sensor, which detects squeezing and its intensity on a continuos scale. The touch sensors at the top and both sides provide information whether the user squeezes the devices from top or from both sides. In addition to that, the ballon also provides an output channel through an air pump that pushes and pulls air bi-directionally. Further, controlling the speed of inflating and deflating allows for different dynamics, which can be used as an additional information channel.

One common application scenario envisaged by the authors involve typical navigation and selection procedures, including scrolling, zooming and similar actions. A particular advantage of the inflatable mouse could be the possibility of applying quick and temporary processes with continuos intensity information, such as quickly zooming in on a map and zooming back to the original state by releasing the pressure on the balloon. But the actual strength of this system is the additional output modality provided by volume-control possibility. The authors propose to render a heartbeat and changing the frequency e.g. in a game to create tension or to shrink the balloon quickly to deliver error messages. Another idea would be to visualise a nap by continuos breathing. An explorative user study revealed that squeezing is suitable for quick tasks but also several problems, such as fatigue after a certain period of usage and the difficulty to maintain a certain depth of squeezing. While some of these problems are of technical nature, the potential of a display with organic look and feel seemed to be promising and has to be further investigated.

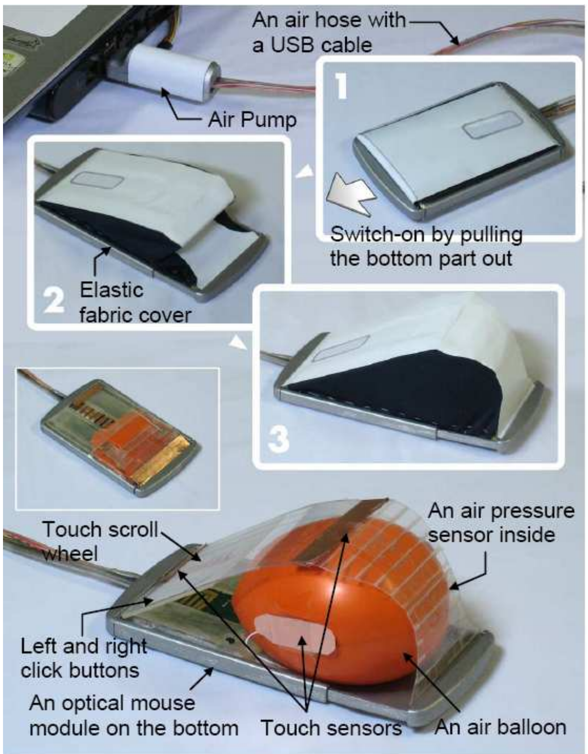


Figure 3.32*: The inflatable mouse (source: [Kim et al., 2008])*

3.3.3.2 Other Inflatable Everyday Items

There are more concepts of inflatable objects that use the changing size of the object as a status indicator similar to ambient displays, but with the notion of life-likeness. One example is Flashbag[5], the inflatable USB stick, a design concept which is not yet being produced. This design series comprises several USB memory sticks adhering to conventional but also animalistic visual styles, which have the ability to grow and shrink according to the relative amount of memory space, which is currently used (see Figure 3.33). This should enable the owner to estimate at a glance whether there still remains enough space for a certain purpose.

Partly similar, the proverbial wallet utilises the haptic channel in different ways to reflect on account balance and previous transactions [Kestner et al., 2009]. Three alternatives of haptic feedback are realised, each named after an animal that the authors associated with the behaviour of that instance. The first wallet example of this ongoing project is called peacock and is able to grow and shrink as well, depending on account balance changes. The general idea here is to promote a sense of financial health and encourage the user to save money. The idea of the mother bear wallet is similar: The item ”protects” the money inside by making it difficult to open, when a certain threshold of a bank account is crossed (see Figure 3.34).

[5] `http://www.plusminus.ru/flashbag.html`, last visited March 17, 2010

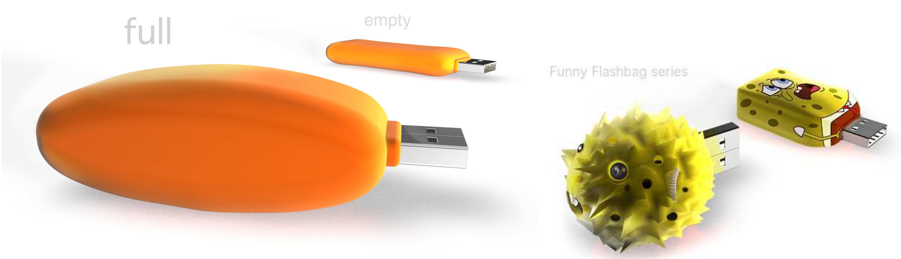


***Figure 3.33**: The inflatable USB memory stick (source: http://www.plusminus.ru/flashbag.html)*

***Figure 3.34**: Prototypes of the proverbial wallet (source: [Kestner et al., 2009])*

The last instance of the proverbial wallet is called bumblebee and simply vibrates whenever a credit card transaction is processed by the bank. This feedback should provide a subtle, physical awareness of the purely virtual process of money deduction after handing out the credit card. All prototypes are still in a proof-of-concept phase with rather bulky instrumentation, the authors were therefore planning to work on more unobtrusive solutions in order to facilitate user studies of this interface.

3.3.3.3 The Shape Changing Mobile

The instrumented mobile phone introduced in this work basically has a plastic board attached to the back of its body that can tilt around a horizontal axis located in the vertical middle of the phone (see Figure 3.35), such that the phone chassis feels either flat or ”thicker” at the top or bottom, depending on the angle of the panel [Hemmert et al., 2010].

The actuation is implemented by a servo motor connected to an Arduino board that communicates over Bluetooth with a nearby PC, which sets the angle of the board. Although this setup would also support an animalistic interface design, the authors concentrated in first studies on conventional feedback and status display using the tactile channel. One use-case covered direct feedback while scrolling horizontally through a picture library holding the phone landscape format, whereby the angle of the panel at the back changes with the position within the sequence. Thus, the more photos remain on one side of the display, the ”thicker” the phone is on that side.

Figure 3.35: *Left: The shape changing phone prototype - Right: A battery status indicator (source: [Hemmert et al., 2010])*

The second application example is the imitation of a download progress bar. Starting with ”thick on top” the back of the phone slowly tilts to the opposite side, reaching ”thick at the bottom” when the download has finished. The last setup is very similar and lets the tilt status reflect the battery status, thus the angle turns as in the previous example from top to bottom, which corresponds to a fully charged battery and an empty battery respectively. A first user study revealed that users adopted quickly to the interaction techniques and were generally able to sense the tilt angles sufficiently. Using a tactile display further had the advantage that the visual channel is still available for other tasks, but at the same time criticism due to a lack of accuracy was raised in direct comparison to a visual alternative. The spatial mapping of the latter example was further reported to be confusing, since there is no clear logical link between start/end position of the tilt and whether it reflects a full or empty battery.

3.3.3.4 Empathetic Living Media

The last example in this section involves bacteria (the so called *E.coli*) as an information carrier that actually lives [Cheok et al., 2008]. By inducing bioluminescence genes the bacteria will glow in the dark (see Figure 3.36). This process involves DNA transformation, which is the uptake of DNA by an organism. The control of this ambient display is achieved with a closed-loop control system which monitors the flow of a control fluid to modify the glow of the bacteria cells. The intensity of the glow is in turn regulated with a feedback unit that measures the glow for feedback purposes, which is necessary since the chemical reactions are not a linear process, thus the amount of induced control liquid is not proportional to the amount of glow that is produced. The visually noticeable elements were the relative light level and changes of light level over time, while in theory it would also be possible to create lights in different colours, which would provide another dimension for encoding information into this display.

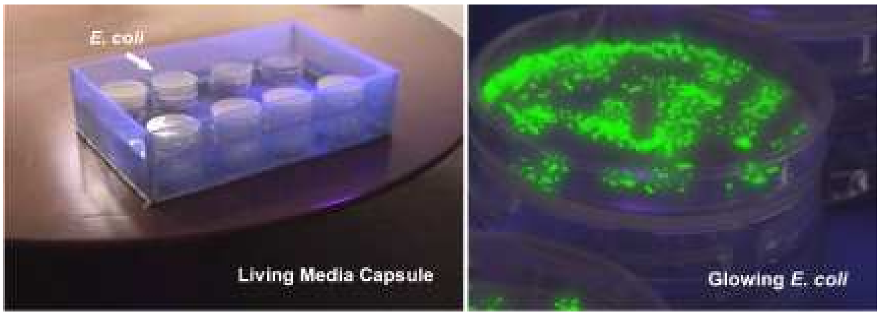


Figure 3.36: *E.coli bacteria as a living media display (source: [Cheok et al., 2008])*

The goal of these research efforts is to evoke emotional reactions that exhibit a different quality compared to the ones provoked by artificial or digital entities. Thus, instead of imitating life, an approach that we have seen in previous examples, living media is exploited and manipulated in order to deliver information to target users. Another major difference is certainly that this work does not employ smart everyday objects as we defined them for this work. But due to its novel and unusual nature, we wanted to include this work into this overview.

The involved research group conducted a user study with the prototypes displaying two different sets of input data: One set corresponded to the amount of chat messages per hour between two friends, while the other one reflected the pollution index of the air in Singapore. These human and environmental domains were deliberately chosen, since the hypothesis of the authors was that such living media is able to achieve higher emotional engagement particularly with social, human, and ecological topics. A virtual instance of the E. coli display was employed as well: A digital version in 3D was rendered inside a window of a desktop computer, showing exactly the same reactions as the real version. The results showed that 87.3% of the users felt more empathy for the real, *living* version of the display. Approximately the same fraction of participants showed a strong acceptance of having the display mapped to human issues and ecological information .

3.3.4 Social Robots

Robots are being developed in research laboratories all around the globe and are expected to be part of our daily lives in near future. Particularly japanese institutes and companies have a strong tradition in robotic research, popular examples are the humanoid robots QRIO by Sony or Honda's Asimo. While much research effort is invested in the improvement of fusion and control of electronics and mechanics and cognitive abilities to enable robots to perform complex tasks, some projects also consider the social component and examine affective impact of human robot interaction. Particularly aspects such as expressing emotion and modelling convincing behaviour are relevant for this work as well, therefore we selected two sophisticated examples of social robots, which we will introduce in this section.

3.3.4.1 The Paro Seal Robot

Paro is a pet-like seal robot (see Figure 3.37), which has been developed by the Intelligent Systems Research Institute of the National Institute of Advanced Industrial Science and Technology in Japan since 1993, and later in 2004 commercialised in collaboration with Intelligent System Co. Ltd. The robot is designed for the elderly to substitute or complement Animal Assistive Therapy and to study the effects of robot therapy (e.g. [Wada and Shibata, 2007] or [Inoue et al., 2008]). Animal therapy is expected to have three effects: Psychological effects (e.g. relaxation, motivation), physiological effects (e.g. improvement of vital signs) and social effects (e.g. stimulation of communication among patients and caregivers). The motivation raises from the difficulties of implementing Animal Assistive Therapy. Most hospitals and nursing homes refuse animals due to potential allergies and infections that could be caused by bites or scratches. In some places, it is not allowed for people to take care of animals in an apartment at all. Apart from that, some elderly who live alone would have difficulties in taking appropriate care of a pet by themselves.

Figure 3.37: *The Paro seal robot (source: http://www.parorobots.com)*

Paro is equipped with four main types of sensors:

1. Light: To distinguish bright from dark environments
2. Audio: For rudimental speech recognition and to determine sound source direction
3. Orientation: To determine whether it is being held by somebody
4. Touch: Tactile sensors all over the surface to measure human contact

Actuators include motors that facilitate vertical and horizontal neck movements, front and rear paddle movements, and independent movement of each eyelid. The robot is also able to render sounds that imitate a real baby harp seal. Several behaviour patterns are implemented to generate a convincing animalistic interaction. Since Paro is able to detect greetings or praise, it can learn to behave in a way that the user prefers. For example, if the robot is caressed after performing a certain action, Paro will remember the previous action and try to repeat it later. Hitting Paro will have the opposite effect and cause it to avoid the previous action. Furthermore, basic emotions such as surprise or happiness are expressed through blinking eyes, or movements of the head and legs.

The behaviour generation system of the robot consists of two hierarchical process layers, which generate three types of behaviour: proactive, reactive, and physiological [Wada and Shibata, 2007] (see Figure 3.38).

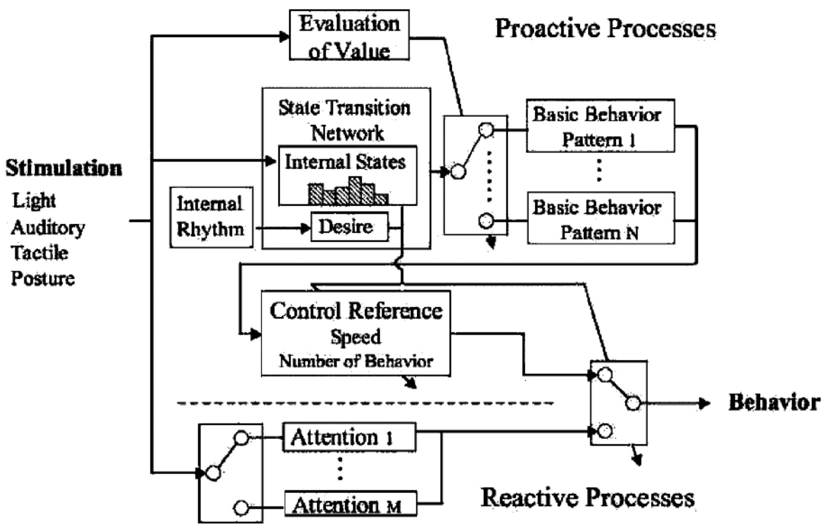


***Figure 3.38**: The behaviour generation system of the Paro seal robot (source: [Wada and Shibata, 2007])*

Proactive Behaviour: Proactive behaviour is based on internal stimulation states, desires, and internal rhythm and is generated by two layers: a behaviour planning layer and a behaviour generation layer. The behaviour planning layer mainly consists of a state transition network and internal states that reflect basic emotions. Each state has a level, which changes according to sensorial stimulation and decays over time. Certain interactions with the robot change internal transition network, therefore defining Paro's *character*. The behaviour planning layer sends basic behavioural patterns to the behaviour generation layer. These patterns include several pre-defined poses and movements. The behaviour generation layer generates low-level control statements that enable the actuators to perform the determined behaviour. The parameters of a control statement depend on the levels of internal states and their combinations. For instance, various parameters can change the speed of actuator movements and create many variations of the same behaviour. Therefore these theoretically almost infinite variations of a finite set of basic patterns are intended to convey a lifelike impression of the robot seal. These parameters are further adjusted according to the priorities of reactive and proactive behaviours based on the levels of the internal states, which makes its behaviour again more unpredictable.

The Paro system further applies Reinforcement Learning to implement a long-term memory: It places positive value on preferred external stimulation, such as stroking and negative value on undesired stimulation, basically beating. Thus, the system can be gradually tuned to the preferred behaviour of its owner. In addition to that, the robot is able to memorise a frequently articulated word as its new name.

Reactive Behaviour: The Paro robot reacts to external stimulation, such as turning to the direction of sudden and loud sound sources. There are several patterns of combining external stimulation with a predefined reaction, which are assumed to be preconditioned and unconscious.

Physiological Behaviour: The system follows a day-night rhythm, which enables the robot to follow spontaneous needs, such as sleep.

To investigate the psychological and social effects of the robots, several user studies in different scales were conducted. A more recent study [Wada and Shibata, 2007] reported on results from tests with elderly residents in a care house, when two robots were introduced into the facility, and activated for over 9 hours each day to interact with the residents. All subjects were interviewed, and their social network was analysed. In addition, the activities of the residents in public areas were recorded by video cameras. For physiological analysis, urine of the residents was obtained and analysed for two hormones[6] This study was the first attempt to investigate the impact of robots such as Paro in situations where subjects can interact freely with the robot. However, the results were obtained from limited number of subjects (12) and a control group did not exist. Summarising the predominantly positive results, it could be shown that the average amount of time spent in a public space of the care house increased from about 1.4 hours to over 2.5 hours. In particular, 1.5 hours of this time included interaction with the Paro robot. Typical situations arose when one person was seen with Paro by another resident, who then stopped to communicate. This time decreased after two weeks, when people were accustomed to the new toy, but still remained on a higher level than before. Further, the values of the hormones significantly improved during the test phase. In particular, significant improvements in the ratio were shown, which was considered as a positive physiological reaction of the residents' organs.

[6] 1) 17-ketosteroid sulfate (17-KS-S), which has high levels in healthy individuals and decreases with failing health or the progress of disease. Also, the 17-KS-S value exhibits sensitivity to changes in psychological and social factors and correlates strongly with a person's will, desire, and energy. 2) 17-hydroxycorticosteroids (17-OHCS), which has high levels in individuals under stress [Selye, 1970].

3.3.4.2 Leonardo

Leonardo is a humanoid, immovable robot with a flexible torso offering 65 degree of freedom. It's outer appearance resembles more an animal like creature than a human and the fact that it is not able to speak [Breazeal et al., 2004] further shifts the perception of this robot toward a more animal-like interaction partner. Nevertheless, it has been designed for social interaction utilising various gestures and facial expressions, particularly for learning/teaching situations in collaborative processes.

Figure 3.39: *Left: Leonardo offers to perform an action. Right: Leonardo asks for help. (source: [Breazeal et al., 2004])*

The overall goal of this work was to improve the intuitiveness, efficiency and enjoyment of human-robot interaction in working and teaching scenarios, by modelling these properties as essentially collaborative processes requiring natural human social skills and conventions. The chosen approach is two-fold, it includes the ability to teach a task to Leonardo through the course of a collaborative dialog with gesture and facial expressions, and the ability to coordinate common intentions to perform a learned task collaboratively.

Cooperative behaviour in this context is considered an ongoing process of *maintaining mutual beliefs, sharing relevant knowledge, coordinating action and demonstrating commitment to the shared activity*. In order to support this, Leonardo employs a variety of gestures and other social cues to continuously communicate its internal state to the human, such as the robot's estimation on who is supposed to do an action or whether a goal has been reached. For example, when then human collaborator has changed a state of the world, the robot acknowledges this action by briefly glancing towards the area of change and then looking at the human partner. This behaviour reassures the user of the robot's awareness of what she or he has done. If this action at the same time fulfils a goal, Leonardo adds a quick confirming nod while redirecting its gaze towards the human. Likewise, nods are used while looking at the human partner to indicate that Leonardo believes that it brought about the completion of a task. This kind of social and reassuring communication is particularly beneficial when the team does not work linearly on a joint plan but instead when each partner works in parallel on different parts of the task, or when unexpected actions of the human occur. Furthermore, the robot has the ability to assess his own capabilities and react accordingly. For instance, if Leonardo would be able to complete a step of the task he will offer to do so, but allow the human partner to override this verbally or by completing it by her- or himself. Accordingly, when Leonardo is not able to carry out an action, it would ask for help. The robot indicates intents to perform an action by pointing to himself and taking up the corresponding pose (see Figure 3.39, left). Similarly, Leonardo expresses the inability to fulfil an assigned task by gesturing towards the human in a help-seeking pose (see Figure 3.39, right). In addition to these gestures, Leonardo shifts its gaze between the human and the object in question to direct the human's attention unambiguously.

The software architecture consists of several modules, comprising speech understanding, vision and attention, cognition and behaviour and motor control. The vision module parses objects from the visual scene, including humans and objects that can be acted on, e.g. buttons. Objects attributes such as colour and location are associated with each perception and forwarded to the cognitive system. Pointing gestures of the human are also recognised and linked to their object referent by spatial reasoning. The continuous stream of perceptions from the vision and speech understanding modules flows into the cognition system where it is integrated into coherent beliefs about objects in the world and their features, e.g. location, colour, ON/OFF state. On top of these processing modules, a set of higher-level capabilities are employed, including goal-based decision making, task learning and task collaboration, which can be reviewed in more detail in [Breazeal et al., 2004].

Leonardo realises several nonverbal behavioural features that are known from graphically embodied conversational agents, but also integrates concepts that are particularly relevant in a physical and spatial context. For example, the posture of its head indicates his field of vision and therefore defines the range of its visual perception. Also, spatially referencing physical objects and the user by gaze and indicating that certain objects are out of reach, are specifically relevant for robots and other entities in the physical realm.

3.3.5 Social Toys

Anthropomorphism has not only been investigated in research contexts, it can also be found in a variety of instances in the commercial arena. The most simple occurrences are toy-like items that playback sound on certain events, e.g. triggered by a remote control, or devices that allow to record personal messages to render such recordings afterwards. For example parrot that repeats twice whatever is recorded, or a cookie jar that sings when opened (see Figure 3.40)[7].

Earlier instances were baby dolls that emanate sounds whenever turned upside down. While the traditional model was purely mechanical, modern variations are electronically

[7] available at http://www.talkingpresents.com or http://www.talkingproducts.co.uk, last visited May 18, 2010

Figure 3.40*: Left: The singing cookie jar - Right: The talking parrot*

driven and are able to distinguish more states than these two. Plastics ducks that squeak when squeezed are also well known and even combine tangible input and anthropomorphic feedback on a very simple level.

Other products have been designed to foster some sort of relationship with their users, often by requiring care taking in order to develop and flourish or by engaging in social forms of interaction. Oftentimes these artefacts also change their behaviour over time or at least provide a highly diverse set of expressions in order to provide a sense of uniqueness or personality. The probably most popular but at the same time the most simple example is the Tamagotchi, a small device with a monochrome low resolution display that shows a creature's lifecycle, starting with hatching from an egg and ending with its death.

Figure 3.41*: Left: Tamagotchi, Right: Furby (sources: http://de.wikipedia.org/wiki/Tamagotchi and http://de.wikipedia.org/wiki/Furby)*

The Tamagotchi requests for the users attention from time to time, who has support basic needs such as food or entertainment, which can be supplied by pushing the appropriate button (see left part of Figure 3.41). Furby is a stuffed animal shown on the right of Figure 3.41, which has been produced by Hasbro. It is equipped with sensors that recognise touching the stomach, back and head and also sounds and movement in the environment. It rudimentarily recognises interaction, for instance if it's being talked to or being caressed. The creature reacts with movements of its eyes, mouth and/or ears. Furthermore, Furby has a voice chip with "furbish" vocabulary, i.e. words that have been invented for this toy. These words are consecutively unlocked on each development stage that are reached by playing with it and feeding it.

***Figure 3.42**: Sony's AIBO (sources: http://commons.wikimedia.org/ and http://www.robotstoreuk.com)*

A much more sophisticated device is Sony's AIBO robot (discontinued), which can be seen in Figure 3.42. AIBO imitates the behaviour of domestic dogs, including walking on four legs, nonverbal communication through ear and tail movements, rolling on the floor etc. The robot also incorporates emotions such as happiness and anger and basic instincts such as the need for companionship. It basically operates autonomously and responses to external stimuli. AIBO can also learn and develop depending on praise and scolding behaviour of the user. For these purposes, the robot-toy is equipped with camera, microphones (including a speech recognition engine) and also touch sensors. Another important aspect of AIBO is its programming interface, which allows to reconfigure and program its behaviour from scratch.

Several years ago, Friedman et al. have examined AIBO related forums and found that 28% of participants reported having an emotional connection to their robot and 26% reported that they considered the robot a family member or companion [Friedman et al., 2003]. It was also noted in the analysis of their studies that AIBO evoked perceptions of life-likeliness and mental states, but it seldom evoked conceptions of moral standing. In this sense, AIBO owners could enjoy their affection on AIBO, feeling a sort of companionship and potentially other psychological benefits, but at the same time decide to ignore it whenever convenient or desirable.

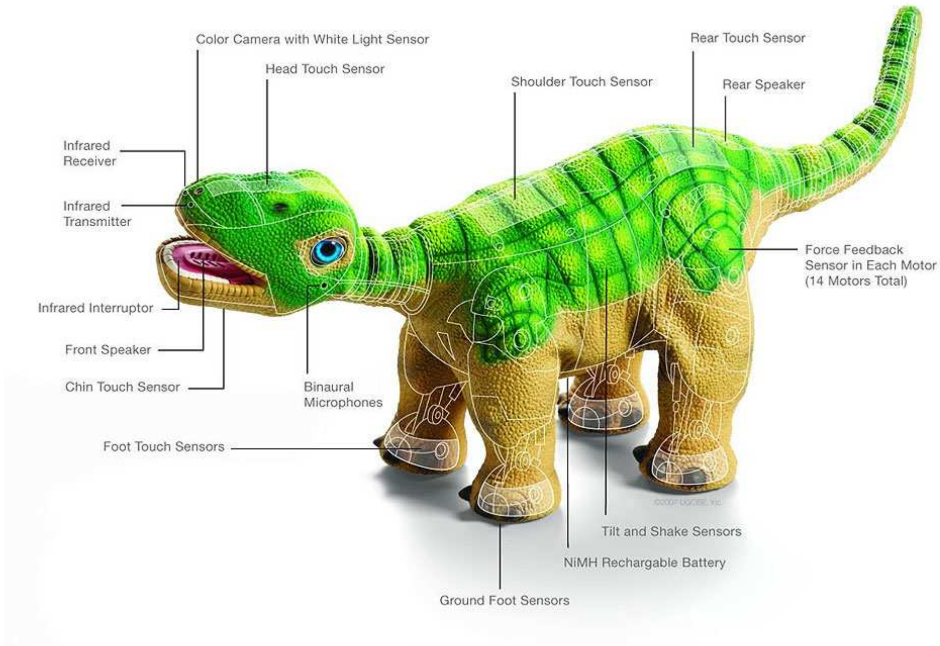


***Figure 3.43**: Placement of Pleo's sensors (source: http://www.pleouk.co.uk)*

The Pleo robot of Innvo Labs (previously manufactured by Ugobe) is another commercial robotic toy that learns and develops over time. This device has the physical form of a young dinosaur and consists of a mechanical frame covered by rubber skin. It is driven by 14 motors, includes two speakers and makes use of a series of sensors, including eight capacitive touch sensors, two infrared distance sensors, a tilt sensor, two microphones and a camera (see Fig. 3.43). One ARM7 32-bit processor acts as the main controller, while another one is responsible for image processing. The default software of Pleo is programmed to develop through the three life stages *hatching*, *infant* and *juvenile*. The first two stages are completed within one hour, resembling its "birth" and first attempts to move and interact with its environment. After that Pleo remains in the juvenile stage, in which it interacts according to its internal *motivational model*.

In an exploratory long term study focusing on everyday interaction with Pleo, Fernaeus et al. have observed that participants were treating Pleo in several ways as if it were a real animal, e.g. by giving it names, petting it and displaying emotions towards it [Fernaeus et al., 2010]. But it has also been shown that these activities generally did not suffice to maintain a long-term interest in the toy, such that after a while it has been put back to the other toys to play with something else.

3.4 Synopsis

In this concluding part we will discuss the relevance and impact of the research activities described in the previous sections of this chapter for the aims of this dissertation. Also, we will summarise which characteristics of each research area are important for this work, and identify the shortcomings of existing approaches and how we will address them with the proposed concept of TASOs. Table 3.1 summarises typical problems and maps these to solutions provided by this work.

The selected studies and prototypes reviewed in this chapter represent a wide field of research spanned by a large number of projects, from which we have attempted to chose an appropriate subset that consists of both widely acknowledged and highly relevant work.

We have seen in several projects how interaction detection can already be realised with rather simplistic technological means, particularly with RFID technology, which allows the recognition of taking or moving a tagged object. On the next level, this passive instrumentation can be extended by active components, such as acceleration sensors, which facilitate much more types of object-centric interaction. At the same time both types of sensors can be utilised for other value-added services as well, e.g. logistics or status surveillance, which is another motivating factor for the development of tangible, object-centric interaction in the UbiComp arena. Looking at the prototypical implementation of a typical sensor-based interfaces, we can often see proprietary hardware solutions, which are tightly coupled with the actual interface. This complicates the reuse of components for further generations of the interface, particularly in these times when available sensor solutions evolve on a fast pace. For this reason we constructed a sensor-based development platform that disconnects hardware solutions from sensor data processing and application logic, which we will introduce in Section 5.1.

The aspect of modularity also relates to another problem in this area: How do we deal with limited sensing and particularly output abilities of instrumented objects? Dealing with instrumented environments, a major component in UbiComp scenarios, allows us to communicate with objects and services in the environment. The architecture of this work integrates environmental components such as sensors and also displays into the design of smart object systems and allows different types of sensor distribution between smart objects and environmental entities (see more on this in Sections 4.2 and 5.2). We have also learned from the Virtual Human project how a modular architecture provides flexibility to developers and an efficient setup of new application prototypes. Certainly, such an architecture is also crucial to support the cooperation of a large amount of partners involved in the software development. We also opt for a modular prototyping architecture (see Section 5.2), basically in order to permit a separation of concerns and to unify distributed components in instrumented environments.

Solutions Offered by the Concept of Tangible Interaction with Anthropomorphic Smart Objects

Problems and Limitations of Related Research	Integrating sensors and services or nearby entities	Integrating and managing environmental displays	Merging concepts of TUI and anthropomorphism	Modular and Reusable Architecture	Anthropomorphic classes with different complexities	Anthropomorphic Auditory Icons	Modeling personality through prosody and vocabulary of speech	Design guidelines for anthropomorphic smart object applications
Limited sensing abilities of instrumented objects	✖			✖				
Limited output capabilities of instrumented objects		✖		✖				
How to adapt smart object interaction to changing environments?				✖				
How to provide intuitive interfaces for quick and casual interaction?			✖					
How can interface design be decoupled from changing hardware?				✖				
How to to deal with reluctance to natural language input in certain contexts?			✖		✖	✖		
How to create customized acoustic feedback that is intuitively understood?						✖		
How to express personality in short term interaction scenarios?							✖	
How to chose appropriate applications for anthropomorphism?					✖			✖

***Table 3.1**: Overview of research problems, limitations and offered solutions*

Some other work like the media control devices has demonstrated the importance of industrial design for the creation of interactive, physical artefacts. The outer appearance of items does not only concern aesthetic properties but also influences the mental model and communicates, for instance, important affordances. We will elaborate this further in the discussion of design guidelines in Section 4.1.5. The I/O brush system, for example, does not only provide the affordances of an conventional paintbrush, indicating how to hold and touch the canvas or other objects, but even imitates the flow of actions of traditional painting and demonstrates how easily an interface can be understood and used when aligned to known interaction processes. In this case the levels of metaphor and embodiment can both be classified as *full*, which is often attempted to be reached by TUI systems in general, as already discussed previously. The concept proposed in this dissertation allows to accomplish both high embodiment and metaphor due to the mergence of anthropomorphism (metaphor) and integrated smart objects (embodiment) into the TUI domain. Furthermore, the technical approach of utilising technology of the environment also supports to shift to other categories of embodiment, if required because of technological constraints such as limited output capabilities of the device itself (more in Section 4.2). In Section 5.3 we also show techniques that help to maintain a high degree of perceived embodiment although output generation is shifted to the environment.

A variety of commercial players have picked up the idea of anthropomorphism in their products, sometimes as part of an overall product concept and sometimes in order to achieve an appealing and consistent impression of a product with autonomous behaviour. This does not only include toys but also tools like the Nabaztag, which does not have any relation to life-likeliness in its original functionality as that of a configurable device for web-related services. The commercial success of some of these anthropomorphised toys such as the Tamagotchi is remarkable. Basically a simple device that represents a virtual living entity, which has to be nurtured by the owner in order to survive and prosper. Similarly, the emotional impact of the AIBO robot on its owners as observed in user forums, confirms the potential qualities of anthropomorphism in the domain of physical products, i.e. providing the possibility of establishing an emotional connection to an object with all its benefits, while users are in fact totally aware that they are only dealing with a technological device.

We have also seen that simple means can already foster an anthropomorphic impression of an artefact. For example, experiences with the shaking toaster and the proverbial wallet revealed that users easily ”recognise” life-likeliness in very simple, one-dimensional displays. We develop and summarise methods for triggering the anthropomorphic stance for smart objects in Section 4.1.4. These aforementioned life-like objects are until now rather isolated solutions without a conceptual framework for applying anthropomorphism in the domain of smart objects. Most of the time, these prototypes are first explorations into the design space to examine single ideas that attempt to invoke some kind of emotional response or empathy at the user. With our work we try to make a first step towards a conceptual framework for the design of interaction with anthropomorphic smart objects.

The results of user studies with the Multimodal Shop Assist have shown that the most preferred combination of input modalities are speech and so called extra-gesture, which corresponds with tangible input in our terminology. This constitutes another strong argument for merging tangible and anthropomorphic user interaction as proposed in this work. The COHIBIT and A.I. Poker installations also underline the potential of this combination. These two systems were build and setup for exhibitions scenarios, a context that involves casual interaction and where exploiting everyday knowledge of users is particularly crucial. For this purpose they combined tangible input with an anthropomorphic interface on RFID basis. But, these two interaction paradigms are yet employed in a disjunct manner, as opposed to the approach of the work at hand. We will discuss further application domains and their appropriateness for tangible, anthropomorphic interfaces in Section 4.1. These considerations also regards the potential reluctance of users to apply natural language input and output in certain contexts, for which we will suggest other, non verbal, means in certain scenarios.

Although human-robot interaction differs in several aspects from interaction with smart objects, both disciplines could inform and learn from each other in certain areas. Specifically interesting for this work is the Paro seal robot, which combines several possibilities of realising believable life-like behaviour. This includes reactive behaviour, which basically consists of performing actions in reaction to external stimulation perceived through sensors, which corresponds to what we call anthropomorphic interpretation of tangible input, elaborated in Section 4.4. Furthermore, physiological behaviour is employed, representing basic needs and desires, as reflected by primitive psychology (see Section 2.4.4). We conceptualise this type of anthropomorphism in an animalistic class of smart objects in Section 4.3.1. Paro's pro-active behaviour resembles a sense of personality reflected by internal states and levels, which also express basic emotions. The state network changes over time through interactions, which generates a high diversity of variations of subtle actuation and thus, creating a sense of life-likeliness through adaption and unpredictability, avoiding ”dumb” repetition. Paro and also other examples introduced in this chapter provide acoustic feedback by animalistic affect bursts, which are easily interpreted and understood. Such anthropomorphic feedback is so far always created for intuition, but for this work we go a step further and conceptualise the generation of custom anthropomorphic auditory icons in Section 6.1.

ECAs often apply unique or at least distinguishable personalities in order to present consistent and convincing characters. Oftentimes this is expressed through behavioural patterns and personality dependent reactions. In certain contexts that are of interest to us, where interaction does not last long, e.g. in shopping scenarios, it will more difficult to deliver a sense of personality with these methods. As a complementing approach we will model personality through prosodic changes and selected vocabulary of speech output, which we will describe in Chapter 8. On the other extreme, we have seen an ECA that is used on a long term over several interaction sessions. The FitTrack prototype implements strategies for relation management, which are for user interfaces still unique and particularly interesting for a certain class of anthropomorphic smart object interfaces. We discuss and adapt these relational concepts to our aims in more detail in Section 4.3.

In Table 3.2 we provide an overview of all projects presented in this chapter and compare them to the work at hand on the basis of key attributes in the realm of TASOs.

After this discussion and analysis of related previous research and projects we will now present the overall concept and design guidelines of our approach in the following chapter.

Based on this we will then continue with describing the implementation of the overall prototyping architecture in Chapter 5 and discuss realisation and evaluation issues in Chapters 8 and 7.

Projects / UI Properties	Smart Object Interface	Retains original function	Involves tangible interaction	Involves anthropomorphism	Merging TUI and AUI	Communicates Affordances	Modular Architecture	Metaphor	Embodiment
TASO	✖	✖	✖	✖	✖	✖	✖	variable, up to “full”	variable, up to “full”
URP	-	n/a	✖	-	n/a	✖	-	noun and verb	nearby
MediaBlocks	-	n/a	✖	-	n/a	✖	-	noun	
I/O Brush	✖	-	✖	-	n/a	✖	-	full	full
Media TUIs	-	n/a	✖	-	n/a	✖	-	verb	environmental
ReacTable	-	n/a	✖	-	n/a	✖	✖	verb	nearby, environmental
Rasa	✖	✖	✖	-	n/a	✖	✖	full	full, nearby
MSA	✖	✖	✖	(✖)	-	-	✖	verb	nearby
Dorfladen	✖	✖	✖	-	n/a	-	-	verb	nearby
COHIBIT	-	n/a	✖	✖	-	✖	✖	noun and verb	environmental
A.I. Poker	✖	✖	✖	✖	-	-	✖	noun and verb	environmental
Clippit	-	n/a	-	✖	n/a	-	n/a	n/a	n/a
Virtual Human	-	n/a	-	✖	n/a	-	✖	n/a	n/a
FitTrack	-	n/a	-	✖	n/a	-	✖	n/a	n/a
Impatient Toaster	✖	✖	✖	✖	-	-	-	verb	full
Thrifty Faucet	✖	✖	-	✖	n/a	-	-	n/a	n/a
Pulsating Phone	-	n/a	-	✖	n/a	-	-	n/a	n/a
Art Examples	✖	✖	-	✖	n/a	-	-	n/a	n/a
Inflatable Mouse	-	n/a	✖	-	n/a	✖	-	verb	full
Shape Changing Mobile	✖	✖	-	-	n/a	-	-	n/a	n/a
Paro	-	n/a	✖	✖	✖	✖	✖	full	full
Leonardo	-	n/a	✖	✖	-	-	✖	noun	full
Talking Products/Toys	✖	✖	-	✖	n/a	-	-	n/a	n/a
Tamagotchi	-	(✖)	✖	✖	-	-	-	none	full
Furby	-	n/a	✖	✖	✖	✖	-	full	full
AIBO	-	n/a	✖	✖	✖	-	✖	full	full

✖ = yes - = no n/a = not applicable

***Table 3.2**: Comparison of related work*

4 Concepts and Guidelines for Designing Tangible Interaction With Anthropomorphic Smart Objects

Life-like characters and anthropomorphic user interfaces in general are commonly applied to various application domains and have been constituting a vivid research area for several years now. We have seen instances of such types of user interfaces in the previous chapter, realised in both purely virtual or material form, such as 2D characters or social robots. Tangible user interfaces have particularly emerged during the previous decade and share a central goal with the aforementioned approach: Build on existing human skills and knowledge to realise more intuitive and enjoyable human computer interaction. We can learn and benefit from previous work in these research areas, but we have to develop new solutions for the particularities of this novel combination of interaction concepts, and identify potential benefits that may arise. The domain of smart object devices further poses domain-specific constraints and opportunities that we have to consider in the formulation of design principles, discussed and analysed in this chapter.

In the following we will develop guidelines for the integrated design of anthropomorphic and TUIs for smart object scenarios, derived from an analysis, combination and aggregation of related research and case studies in the respective fields [Schmitz, 2010] [Schmitz, 2011]. Thus we are attempting to support developers and designers of smart object systems in deciding which level of anthropomorphism is adequate for the intended application scenario and by which means it can be achieved.

We will start with design rules to help with basic design decisions of different kinds, followed by technical considerations regarding different approaches of instrumentation. Furthermore, we will introduce different classes of anthropomorphic smart objects and describe possible interpretations of tangible input in such anthropomorphic scenarios and provide guidelines for the selection of sensors for object integrated input recognition. We also describe novel approaches of expressing affect and personality in acoustic interfaces of TASOs.

4.1 Design Rules

Interactive systems design can have plenty of different goals and intentions, depending on functionality, application domain, target users, budgetary constraints and an abundance of other parameters. However, two important objectives are common to most systems: A high degree of ease of use and joy of use, as previously discussed in Section 2.2.1.1. Both concepts describe essential aspects in terms of interaction efficiency and enjoyment, respectively. Hedonic qualities might have different levels of importance in pro-portion to ease of use, depending on the purpose of the system the designers' aims and further interests of other stakeholders. Since social factors are normally beyond reach of the designer, we will first discuss the applicability of anthropomorphic and tangible user interaction to different types of systems and tasks.

4.1.1 Mental Model of Anthropomorphic Objects

The choice of a certain interaction paradigm always influences what is often referred to as the *mental model* of the interaction or functionality of a smart object. A mental model describes what a user thinks how a process or system works. It represents a logic of relationships between various parts and a concept of action consequences. Consequently, a consistent mental model is essential for intuitive interaction, in which the user feels in control - aspects that impact the joy of use (see Section 2.2.1.1). Many aspects influence the mental model, such as affordances and system feedback. Applying an anthropomorphic paradigm strongly builds certain expectations towards system reactions and thus fundamentally interrelates with the mental model. The importance of a *balanced* anthropomorphism has been noted oftentimes (e.g. [Duffy, 2003] [Shneiderman and Maes, 1997]), saying that anthropomorphic features should hint at certain capabilities that meet our expectations, or possibly even surprise and surpass these expectations.

First of all, anthropomorphic smart objects will typically be perceived as **independent actors**, causing the attention of the user to focus on the physical object as the main interaction partner. This presumes that the overall interaction logic matches the user's object centric perception to avoid confusion or an inconsistent mental model. In a concept that would employ an artefact as mere carrier (as in e.g. Section 3.2.1.3) of data for later inspection through further systems or tools, the application of a meaningful anthropomorphic representation to this artefact will be more difficult. Another inapt application might leverage smart objects as tools that are used to operate a larger system, as the building models in the URP system [Underkoffler and Ishii, 1999] introduced in Section 3.2.1.2. In these cases, the idea of anthropomorphic objects might potentially conflict with the overall mental model, since the interactive objects act as tools or data holders and therefore constitute dependent parts of a larger whole. Certainly, these borders between independent actors, tools and information carrier can be blurred, depending on the actual application.

Thus the object centric mode of interaction is facilitating anthropomorphic interaction - a thesis also strongly supported by K. P. Fishkin in his taxonomy introduced in Section

2.3.1: He argues that a high degree of embodiment is helpful for the perceived tangibility of the interface, which reduces learning efforts, increases the power of the metaphor and therefore adds to the value of the application. In general, a strong embodiment is achieved when the state of computing is perceived as inside the interactive device, which again can be accomplished by placing in- and output at the artefact itself (more about technical integration in Section 4.2).

Work by Wasinger and Wahlster has revealed that users dislike to directly converse in spoken dialogues with a bar of soap but would be generally willing to talk to digital cameras, cars or personal computers [Wasinger and Wahlster, 2005]. This suggests that speech as a sophisticated interaction modality might only feel natural when applied to a physical object, if the object is perceived as an item with inherent complexity, such as electronic appliances or computer systems. One reason for this might be that it feels awkward for the user to build up a mental model, in which a very primitive object is capable of managing rather complex computational tasks such as speech processing. Instead, a more complex and compound artefact, the inner workings of which are not completely grasped by the user, might be more suitable for the attribution of such abilities. These findings might be taken as guidelines in the choice of application scenarios, suggesting to select objects, which do not appear to be obviously incapable of computational functions at all, such as the aforementioned bar of soap, stones, food etc. It could be speculated, though, that a particularly designed shape of an otherwise simple object might counteract this phenomenon (see also Section 4.1.5). However, by introducing simplified means of anthropomorphic tangible interaction (see more in Section 4.3.1) we provide a first step in providing a rather simplistic anthropomorphic metaphor in order to be able to apply it to relatively primitive objects as well.

4.1.2 Level of Professionalism

Another important aspect of the design space is that of the degree of required *professionalism*, as we call it. This professionalism affects the trade-offs that may emerge between the goals of joy of use and ease of use as described in Section 2.2.1. For this work, the term professionalism comprises the frequency of (re-)use of a system or the repetitiveness of a task and the required accuracy and reliability. In this sense, highly professional systems are therefore be utilised by trained expert users and applied on a frequent basis, demanding accurate results and potentially a high degree of safety. Intuitive interaction does not necessarily have a high priority in the design of such systems, since training of users can be expected to achieve a higher overall efficiency in the long run, since the learning costs in such contexts are still in positive correlation to longterm efficiency and time-saving. Examples of this task domain include applications in safety-critical working environments, such as in airplane controls or emergency systems in a nuclear power plant.

On the contrary, novice users and users that only casually interact with the system often cannot be trained, simply due to time constraints in the interaction context. Typical examples for such scenarios are exhibitions, shopping environments, rental cars, hotel rooms, airports and airplanes where visitors or users might be exposed to a multitude of interactive systems (i.e. products or physical installations) without the possibility of learning a complex interaction process. Aiming at such scenarios with anthropomorphic smart objects is also supported by Preece et al., who have stated that conversational user interfaces encourage especially novices and technophobes to interact with the system in familiar way that makes them *”feel comfortable, at ease and less scared”* [Preece et al., 2002]

But even without tight time constraints a tedious learning curve is often not appropriate and could lead to neglecting the system in certain contexts. In such particular situations, the benefits of both anthropomorphic and TUIs can play out well. Furthermore, exploratory applications that aim at entertainment or edutainment have shown in various examples that they may benefit from anthropomorphic as well as from tangible interfaces, e.g. the Virtual Constructor (see Section 3.2.3.1) and A.I. Poker (see Section 3.2.3.2).

Additionally, anthropomorphic tangible interfaces have the potential to narrow the gap between novice and expert users, if offered as an optional interaction style, which can be replaced on demand by an interface alternative more suitable to the particular demands of an expert user. For example, interactive products in a supermarket could provide different interfaces for customers (the typical casual users) than for employees of the store (expert users), who will in most cases carry out different tasks in the same supermarket context.

Another class of smart object applications on the scales of repetitiveness and accuracy involves systems that are used on a regular, but casual level by novice users. Examples for such scenarios comprise home and kitchen appliances (as described in e.g. Sections 3.3.2.1 and 3.3.2.2) that provide uncritical functionality that are accessed frequently over a longer period of time. We believe that a high degree of enjoyment and intuitiveness is specifically appropriate in a home context, minimising the need for learning and memorising smart object interaction in this casual environment. Still, the possibility of emerging fatigue or annoyance as a result of a high rate of repetition of the interaction has to be taken into account with the overall design of the system.

4.1.3 Extending Traditional Use

Smart objects are, according to our definition in Section 2.1.3, items that already exist as traditional, probably purely analogue objects with their respective functionality in the original, physical context. Therefore, additional services that are enabled by technological enhancement of the original object will automatically be put in relation to the original use of the item with the potential of creating conflicts in several dimensions. In this section we will identify and discuss three design dimensions that are important for the creation of anthropomorphic smart objects with respect to the original use of the artefact: **Function, Interaction and Modality**. Chi et al. [Chi et al., 2007] have analysed smart object systems designed for living room environments and formulated design heuristics that regard the functional relation and the interaction relation between the digital and the traditional usage. We can partially adapt and transfer these heuristics into our design considerations.

4.1.3.1 Extending Functionality

The functionality added to an object can be directly related to the object's original purpose, thus *enhancing* it. For example, a flowerpot that detects and communicates water shortage would directly enhance its traditional purpose of accommodating a plant. A weaker relationship to the original nature of an object could still *complement* the physical functionality and, for instance, provide services that are useful in the same context and domain. A completely *unrelated* functionality would appear as totally independent from the physical item and its purposes. Certainly, boundaries are not always clear and at times could be crossed by objects that provide several services with different kinds of functional relations to their traditional purposes. For instance the TeCo Mediacup (as introduced in Section 3.1.1) prompts the coffee machine to boil new coffee when the cup is almost empty (enhancing) and also sends meeting-notifications to the doorplate when several of these cups are present in one room (unrelated). Although unrelated functions may make sense in certain scenarios, we believe that a strong relation to the original function is beneficial due to several reasons: First, it makes it easier for the user to associate digital functions with an object - particularly in environments with a potentially high amount of smart objects. Second, the probability that the object is available when it is needed is higher when its original use relates to the situation at hand. Third, the interaction and product designers' task of creating a device with the appropriate affordances for both physical and digital functionality might be unnecessarily more complicated when alienating an object from its original purpose.

4.1.3.2 Extending Interaction

The dimension of interaction relation can be characterised similarly: A *natural* interaction method utilises the original method of use, disburdening the user of learning any new interaction styles, as for instance seen in the I/O brush (as introduced in Section 3.2.1.5). An *intuitive* use is different to the physical one, but strongly relates to it, which greatly simplifies the learning efforts of the user. Consequently, an *unrelated* interaction method has no resemblance to the object's original interaction. We basically agree with the assessment of Chi et al. that smart object interfaces with a *natural* or *intuitive* interaction relation potentially yield interfaces that are easy and quick to learn, but we argue that applying alternative interaction paradigms can nevertheless be a beneficial solution (such as anthropomorphism in smart object systems), as long as the overall goals of usability and user experience can be met, building on the user's needs and abilities. Novel ways of interaction may possibly even foster a new interest in otherwise unattractive devices, resulting in additional inspiration and motivation. But, it is important for the designer to keep in mind that a low correspondency between the original and new interaction method requires more attention in communicating the interaction possibilities, e.g. through affordances (see Section 2.3.2).

4.1.3.3 Extending Modalities

This question also leads to the choice of the modalities for interaction realisation. The concept of symmetric modality as introduced in [Wahlster, 2003] suggests that multimodal systems should provide the same input methods as used for output and vice versa. Since TASOs inherently employ physical input as a modality, symmetric modality suggests the use of haptics for the output channel. From a technical perspective, TASOs as potential multimodal systems also have the advantage to reduce the overall uncertainty in user input recognition through mutual disambiguation of various analysis results, as described in e.g. [Oviatt, 1999]. By fusing symbolic and raw sensor data derived from recognition components, certain recognition errors can be avoided such that the overall recognition performance is improved.

Anthropomorphism can be established through speech and appropriate sound (see Section 4.3). We consider the combination of the haptic and acoustic channels as primary interaction means for anthropomorphic smart objects, in combination with tangible input due to the natural tangibility of physical objects per se. As opposed to the acoustic channel, which could be delivered by invisibly integrated speakers, the visual channel might in many cases not be available, since we focus on smart objects, which are designed for an analogue use with physical constraints and requirements, making it difficult to alter shape and outer appearance with display technologies. In future, technologies such as OLEDs (organic light emitting diodes) will enable developers to integrate flexible displays into certain types of objects, which would for instance allow for displaying graphically embodied agents. This will provide additional means of expressing life-likeliness in smart objects.

Obviously, the choice of input and output modalities of the smart object interaction and services interacts with the modalities utilised by the object in its original use. In accordance with the observations made by Chi et al. [Chi et al., 2007], the nature of this relation can generally be conflicting or enhancing, or the smart object's interaction modalities might be completely unrelated to the original ones. A conflicting situation might occur when the original object's functions make use of the same channel that is used by the digital extension. For example, a radio that is playing music and at the same time implementing natural language dialogues will have to mix speech output into the audio stream, which might already include speech and thus potentially confuse the user. Not to mention the technical problems that arise for the speech recogniser in noisy environments. But, the implications of this example cannot be simply generalised from the auditory channel to other modalities, instead each modality and basically each interaction concept has to be analysed independently. The use of the haptic channel, for instance, even benefits from requiring tangible input, since touch is necessary to deliver haptic information. On the other hand, if the original use of the object would involve strong haptic forces as in squeezing or pushing (e.g with tools like hammers or screwdrivers), additional haptic feedback might not be perceived by the user and the recognition of tangible input might get difficult as well. Modality enhancement can be achieved in interaction situations, where the user is already accustomed to observe a certain modality channel, which can be exploited to establish new interaction methods. For example, some home appliances use simple synthesised or mechanical sounds to reflect system status or provide user feedback, like toasters, water heaters, dish washing machines. We also include feedback sounds into that category, which are not intentionally created for such purposes but still utilised by the user to infer a system state, such as the intensive sound of a spinning washing machine in its last washing phase. Thus, the relation of original and additionally employed modalities poses both restrictions and potentials to the interaction designer, which have to be regarded carefully as the case arises.

4.1.3.4 General Output Characteristics

In addition to that, the general characteristics of the available output modalities have to be considered as well: The haptic channel provides a very unobtrusive communication modality, which makes it adequate for preserving privacy or for contexts that do not allow acoustic output. However, synthetic tactile output such as vibration has a rather low resolution compared to acoustic or visual displays and therefore qualifies better for low-attention monitoring [MacLean, 2000a] as for instance applied in the shape changing phone or the *Flashbag* introduced in Section 3.3.3. Furthermore, vibration and similar haptics provide simple means to indicate the presence of a life-like entity. As already mentioned, audio might be an inappropriate medium in certain environments and exacerbate acoustic pollution. But it can also be advantageous to be able to perceive audio without dedicating explicit attention to the source. As a side effect in situations such as exhibitions, sounds might also spark interest of collocated and previously unaware visitors. Moreover, using the haptic or acoustic modality has the important property that it relieves the visual channel, which is generally the primary human sense and main source of incoming information.

4.1.4 Triggering the Anthropomorphic Stance

Applying anthropomorphism changes the way how users try to understand and make sense of a computer by projecting everyday expectations of human and animalistic behaviour onto it. An anthropomorphic interaction paradigm has the potential to apply a strong metaphor of verb, which regards to the motion of an object or to the interaction in relation to the real-world action [Fishkin, 2004]. As we could see in Section 3.3, the anthropomorphic stance can be triggered by various means, which we will summarise and adapt for the domain of smart objects in this section.

Humans are inherently well-trained to perceive life, already from childhood on [Scassellati, 2000]. The device's **visual appearance** will often shape first impressions and can therefore play a considerable role in establishing first expectations. Organic shapes or those that resemble limbs, parts of the face and other body parts can be effective visual cues (see also Section 4.1.5), which has also been applied by systems described in [Horvitz et al., 1998] [Wada and Shibata, 2007] [Breazeal et al., 2004] and social robots and toys as in Sections 3.3.4 and 3.3.5.

The mere **presence of voice** is another strong trigger for anthropomorphic perception [Persson et al., 2000], irrespective of contents and sounds. The effect can be intensified by increasing its naturalness (see also Section 4.3.2) as done by work such as [Göbel et al., 2004] [Ndiaye et al., 2005] or [Gebhard et al., 2008]. Besides speech, the acoustic channel can be used to deliver alternative feedback (compare also with Section 4.3.1) that resembles **life-like, physiological sounds or haptics** instead of conventional sound feedback, for example pulse, breath, snore or cough, which is for instance applied in [Hemmert and Joost, 2008] and [Wada and Shibata, 2007].

Furthermore, **pro-active and autonomous behaviour** can be a fundamental strategy in order to evoke anthropomorphic interpretation [Persson et al., 2000]. This can already be instantiated with very simple means, for example by autonomous movement, which is the most basic distinction of living matter [White, 1995] and for instance applied in [Burneleit et al., 2009] [Togler et al., 2009]). On a next level, **stereotypical behaviour**, which expresses e.g. shyness or curiosity, addresses more complex processes but adds essentially to anthropomorphic strategies [Hoffman and Breazeal, 2007] [Wada and Shibata, 2007] (see also 3.3.2.2 for the (In)Security Camera and the Luxalive Lamp as examples).

Employing memories about previous experiences, such that they alter or influence the current state and behaviour, can be another effective method to display consistent life-like entities. This is particularly useful for smart objects that already have the sensory means to record aspects of their life-cycle, which makes them inherently unique and individual, strengthening the impression of a social being (see also [Duffy, 2003]). This can be further supported by not only considering actual interaction sessions with the user but also the remaining period of time, during which the object persistently ”exists” independently from the user. That means in consequence that a TASO is a computer individual, which is never turned off, although it might ”rest”, ”sleep” or just ”wait”.

Applying anthropomorphic concepts into different kinds of smart object scenarios will be elaborated in more detail in Section 4.3.1.

4.1.5 Industrial Design Challenges

A product designers traditional task is to design the outer appearance of an object, generally aiming at an aesthetic visual design. The latest generation of products, with embedded electronic components and richer functionality, extends the conventional scope to the design and the composition of interactivity [Butz et al., 2005]. In this section we will examine the consequences the particular characteristics of anthropomorphic smart object design have for industrial designers, starting with general considerations in the design process.

As defined in Section 2.1.3, we consider smart objects as objects that already exist in their original, analogue world with certain purposes and requirements that specify or constrain their general shape. Certainly, these traditional requirements still exist and constitute primary parameters. Integrating electronic components is a common part of the design phase in industrial design, such that the additional instrumentation of a smart object is not a completely new process. But, the more complex requirements of sensors and actuators and their interrelation with other design parameters including materials, shape and mechanics pose new challenges to the stages of both design and construction. The principle of prototypical realisations is not very different to creating a product for large scale production, except that mistakes in prototypes are certainly more forgiving and less expensive. A prototype, though, has to meet different requirements that are often very different from or even contrary to design expectations towards a final product, such as flexibility in hardware rearrangements or easy access of interiors. Therefore, prototypical electronics would normally be redesigned from scratch and tailored to the the final product, also to optimise costs.

Within the visual scope for design, the outer appearance can have an important impact on the overall perception of an anthropomorphic smart object (see also Section 4.1.4). The concept of the metaphor of noun (as discussed in Section 2.3.1) regards the analogy of the shape, texture, look and sound of objects to their metaphorical counterpart. A high analogy is therefore obviously advantageous and can be achieved on the visual channel by fostering a first primitive categorization of living matter. We can facilitate an anthropomorphic stance towards the object by applying an organic shape or by resembling configurations that look like faces and bodies (see examples in Figure 4.1).

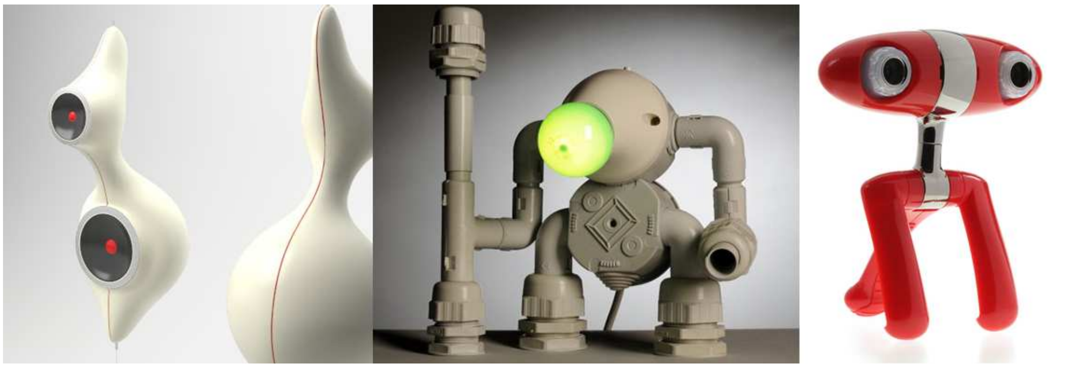


Figure 4.1: *Anthropomorphic product design examples, resembling limbs and facial elements. (sources: http://www.madebymakers.dk (Loudspeaker), http://www.robolamps.eu (Lamp), http://www.minoru3d.com (Webcam))*

Also limbs or just facial elements have a particular salience to humans and help to express anthropomorphism [Persson et al., 2000]. The texture of living objects could for instance be imitated by soft and furry materials at the outmost layer. It is not necessary to overly exert anthropomorphic product design, small efforts can already lead to good results [Reeves and Nass, 1996]. It has further been shown that the overall impression of appearance triggers certain feelings and action tendencies, and thus, a general stance towards an entity, which can be utilised in the design process as well. For instance, ”cuteness” triggers feelings of ”infantile helplessness” and a tendency to care for and protect. Accordingly, perception of ”ugliness” and ”beauty” activate an action tendency to avoid or to approach that particular face or body [Tan and Fasting, 1996].

As argued in Section 4.1.1, people tend to refuse to directly address simple objects in natural language interaction. To counteract this phenomenon the application of more complex shapes might help to attain a less simple impression of the object on a visual level, and thus to create to a certain extent the general impression of a more complex device.

Other techniques and metaphors might further help to overcome limitations that originate from inappropriate appearance: For instance, a genie-in-a-bottle effect could be applied by creating a virtual anthropomorphic actor that is associated with a device. Such a genie effect, leading to the user perceiving the virtual character as representing the physical device, could be realised for example by a (projected) animation, indicating that the character *morphs* out of the object.

Affordances (see Section 2.3.2) are also an important factor in the physical design of anthropomorphic smart objects and interrelate with what we will call **anthropomorphic affordances**. This notion describes affordances that arise from an anthropomorphic stance and the impact on shape and its affordances. Furthermore, affordances can also be utilised to suggest interaction modalities: Body parts that are associated with senses could indicate that the object is capable of perceiving the corresponding modality (e.g. eyes suggest seeing and ears hearing, respectively). The resemblance of limbs, facial components or other body parts afford new interaction possibilities that are beyond the traditional affordance characteristics, which are mainly based on physical and mechanical laws. Common examples are the big belly of a chinese buddha that invites to rubbing, while a horizontally stretched arm might invite for shaking the hand. So, if the user has already taken up the anthropomorphic stance, a yet unexplored type of affordances comes into play that has to be analysed by the designer as part of a holistic design process.

4.2 Instrumenting Objects and Environments

One major design decision on a technical level regards the degree of instrumentation of the smart object itself and its environment. This is not a particular issue of anthropomorphic smart objects but basically of all applications that combine smart objects and intelligent environments. In this section we will discuss the possibilities and implications of embedding technology inside the object itself and shifting it to the environment, specifically for prototyping anthropomorphic smart objects.

For the purposes of this work we distinguish four different classes of smart object implementations, which are also visualised in Figure 4.2:

1. **Self contained**: All instrumentation resides inside the object.

   A self contained smart object is able to perform all interaction methods and services without any other support of the environment or other infrastructural elements besides electricity supply, either continuously or through occasional charges. By integrating sensors, actors, computation and communication capabilities within the device, the requirements of cyber-physical systems as introduced in Section 2.1.1 are fulfilled as

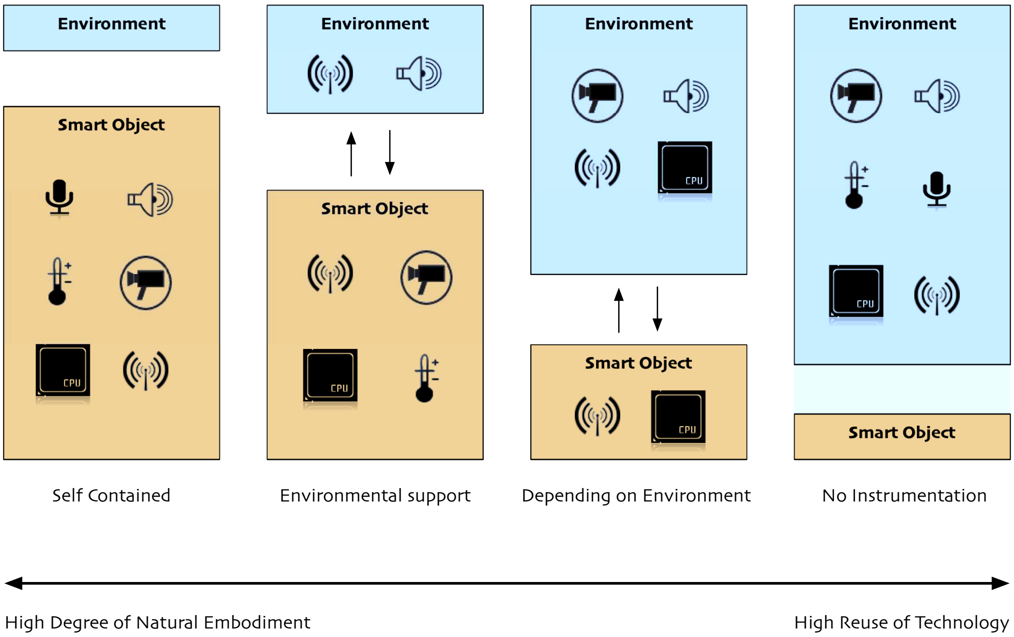


Figure 4.2: *Different classes of integrating technology in smart object systems and instrumented environments*

well. Obviously, a major advantage is the autonomy of the device and thereby the possibility to easily transfer the system to any other location. The mobility criterion is much more critical for products and real-life applications than for pure prototypical development in laboratory environments or similar scenarios. However, the integration of all necessary hardware poses strong constraints on the overall physical construction of the prototype, depending on the application and its requirements. Each component requires space and adds weight, and many sensors and actuators require careful placement to accomplish their task. But at the same time, such a solution supports an object centric interaction model and yields *full embodiment* according to Fishkin's taxonomy (see Section 2.3.1).

2. **Instrumented with environmental support**: Primary technology, which is required for main functionality, is embedded into the object, the environment provides additional services.

   Instrumented smart objects with environmental support as well are able to function independently retaining their basic functionality. But, in addition to that, such objects are also able to communicate with components of the environment, which possibly support the object with additional services. For example, cameras might identify the interacting user or loudspeakers render audio output. The main advantage of this approach is the possibility to utilise theoretically unlimited resources of the surrounding infrastructure and thus realising otherwise unfeasible functionality. It might also be advantageous to outsource hardware components and/or functionality to the environment in order to save space, reduce the object's complexity, and to make externalised components reusable for other smart objects. Of course the mobility advantage diminishes with increasing dependency on surrounding infrastructure. Regarding the mental model of an object centric interaction, the developer has to consider the possibility that shifting sensors, computation or actuators outside the object may alter the users perception and conflict with the goal of achieving a high degree of embodiment. But this issue also depends on the employed modality and how contents are delivered. For example, auditory output could in many cases still be perceived as emitted by the device if rendered appropriately, e.g. through spatial audio (see more in Section 5.3.3). The result on the scale of embodiment would most likely be *nearby* or *environmental*.

3. **Partially instrumented with main technology in the environment**: The object is equipped with a limited amount of instrumentation (e.g. RFID tag) and the environment contains further required technology (RFID reader, sensors).

   A partially instrumented smart object that relies on the availability of certain infrastructural technology goes one step further and necessitates a specific environment in order to function as intended. An example would be an object that relies on an external computing machine to process and interpret sensor data, required for the interaction process. The apparent disadvantage of the dependency on technological infrastructure is a trade-off to the pragmatics of a centralised solution that simplifies development efforts of a multitude of prototypes e.g. due to fast changing requirements. Limited hardware resources can be efficiently utilised, shared and reused. On the other hand, choosing this approach foregoes obtaining insights on the technical feasibility of a higher level of embedment in real-world scenarios. Apart from that, certain sensory data requires to be collected within the object itself, e.g. for tangible input detection. Analogical to the previous class of objects, sustaining the perception of a self-contained and highly embodied object as the main interaction partner in this case is more difficult and needs to be addressed by other means.

4. **No instrumentation**: All technological components reside in the environment.

   At the other end of the scale we have smart objects that do not make use of embedded technology at all, but instead are part of an instrumented environment that hosts all application logic, sensors and display devices involved in the application. This is the case for instance when cameras are used to identify and track an object, while external displays are responsible for the system output. This situation exposes the developer to several problems:

   - Interaction recognition cannot make use of integrated sensors (e.g. for acceleration)
   - A strong dependency on a very particular instrumented environment exists
   - Creating an impression of high embodiment is more difficult. More likely, the perceived embodiment would be classified as *environmental* or *distant*

A positive side effect of working with devices of this class of instrumentation is that the original object design can be adopted, relieving product designers from the burden of integrating additional technology.

4.3 Anthropomorphic Smart Object Classes

Aligned to Persson's analysis of multi-layered anthropomorphism, which we reviewed in Section 2.4.4, we introduce three general types and in total five distinct classes of anthropomorphic smart objects. We propose three basic types: Animalistic Objects (yellow), Talking Objects (blue) and Bonding Objects (purple). Animalistic Objects are further split into Basic and Social Animalistic Objects, and Talking Objects into Talking Objects with Personality and Emotional Talking Objects. For defining these classes we distinguish levels of anthropomorphic perception on one axis and attributes of software agents regarding their autonomic behaviour introduced in Section 2.4.3 on the other axis (see Figure 4.3). The selected agent capabilities directly relate to an anthropomorphic interpretation of systems and potentially support the realisation of anthropomorphic systems. Thus, the graph outlines the relation between the classes of anthropomorphic smart objects, the anthropomorphic layer they build upon, and different types of system behaviour from the perspective of software agent technologies. Different levels in each of the two dimensions imply that lower level concepts are subsumed, which means, for instance, that the anthropomorphic layer of personality includes folk psychological phenomena, which constitute more sophisticated concepts of emotion than those associated with primitive psychology. Accordingly, mentalistic capabilities of agents generally require autonomy and pro-activity, which are extensions of simple reactive agent systems. These classes facilitate the categorisation of previous and future work, while the boundaries of each class are not necessarily fixed limits. Instead they suggest directives for modelling anthropomorphic smart objects consistently.

We have further applied these categories to selected projects that were introduced in Chapter 3 as much as possible even though some examples are purely virtual. Some classifications were difficult and details made the difference, as for instance with the projects COHIBIT and Ideas4Games. COHIBIT involves VCs that incorporate emotional responses to user actions, while the IDEAS4Games installation focuses not only on emotional believability but also on distinct, stereotypical character profiles of the two virtual poker players. Further, the Leonardo robot realises continuity and social behaviour that are beyond animalistic limitations, but at the same time it is not able to speak and its visual appearance resembles an animal-like character, which makes an unambiguous classification difficult.

We complement the design rules of this chapter with more specific concepts about variations of tangible, anthropomorphic smart objects. For each class we start with a general description of their relevant characteristics, propose guidelines for their realisation and outline suitable application scenarios.

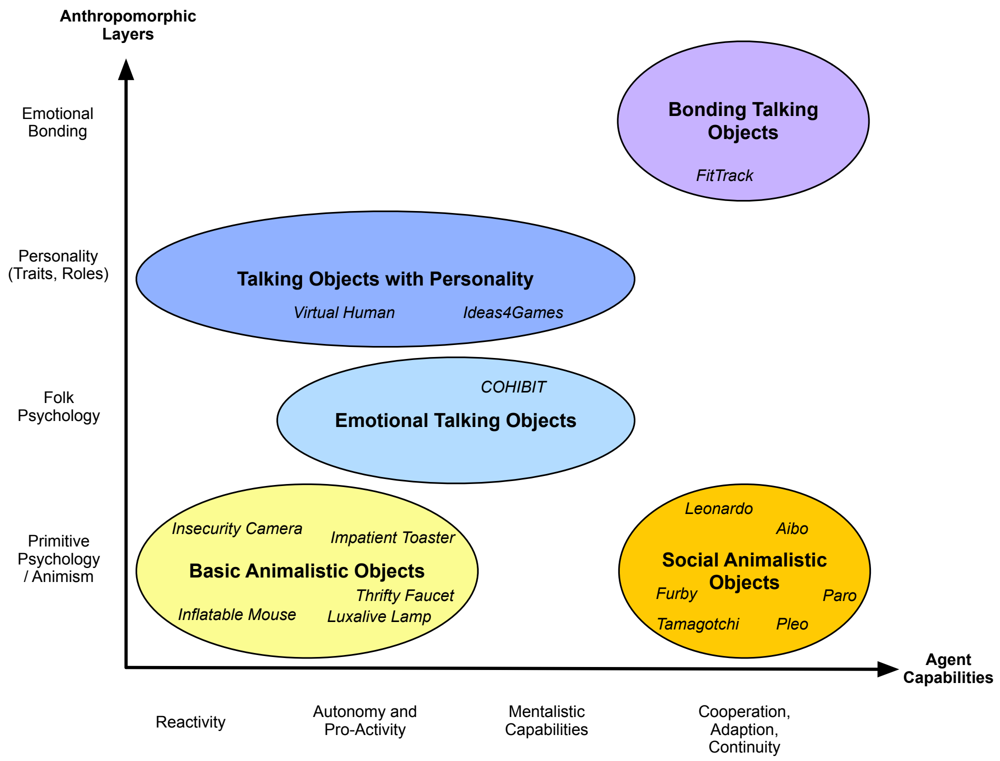


Figure 4.3: *Relation of anthropomorphic smart object classes*

4.3.1 Animalistic Objects

Basic desires can easily be projected onto simple behaviour, exploiting expectations about such needs and desires. This includes actions that relate to life preservation, pain, hunger, desire for sleep, exhaustion, fear, excitement and similar conditions. This relates to the concept of *Primitive Psychology* and allows to predict and interpret simple animalistic behaviour immediately [Persson et al., 2002].

Basic Animalistic Objects can be realised by mapping important system states to suitable basic needs and implementing reactions to status changes by associated animalistic expressions. For example, an object that observes its state and attempts to preserve primary conditions (e.g. intactness, energy level, temperature) could render acoustic animalistic feedback if one the conditions is violated. For instance a sound of pain, when the object falls, is being squeezed or has detected damage by other means. Sounds of fatigue or exhaustion could be mapped to low energy status and inform the user pro-actively. If the temperature exceeds or falls below a threshold, the reaction could consist of moaning sounds or shivering. If the mapping is well-thought-out and the critical parameters of the object known by the user, an understanding of the anthropomorphic feedback will happen intuitively. If the user is then taking action to correct the issue (e.g. recharging the object or taking it to a cooler place), the corresponding feedback can again inform about the state change. The tangible component in this class of objects is important as well, as it fits to the immediacy of this anthropomorphic class. Touching the object can already trigger a response to let the user know that the object recognises and values that its issues are being addressed - this would not only provide intermediate feedback, but also intensify the anthropomorphic stance. In general, tangible input as described in Section 4.4 also applies well here.

This type of animalistic smart objects can be applied to comparably simple artefacts, which employ functionality that can be modelled as action/reaction systems depending on given parameters. This also avoids the problem that people refuse to talk to simple objects [Wasinger and Wahlster, 2005] by providing simplified anthropomorphic means, and thus an appropriate mental model. An example for this class of smart objects might be the smart pizza packing [Schneider and Kröner, 2008], which observes its own state and checks whether, for instance, the cool-chain is interrupted. Other examples can be imagined in domestic environments, e.g. electronic home appliances. A coffee machine would appear sad, when it is supposed to prepare coffee but has to be refilled with water or coffee powder before. Or annoyed, when the brewed coffee is not picked up for some time. A USB memory stick could be shaken to check the amount of stored data, the feedback would resemble nauseous sounds in different frequencies and speed to reflect the amount of space taken up by data with a low frequency of clunky, dull sounds correlating to a rather full state and high frequency of high, agile tones correlating to an empty state.

If necessary, this sort of animalistic interaction paradigm can be extended by conventional means of displaying more detailed information. For instance, a temperature drop below a certain threshold might be followed by an inquiry about the exact temperature or for how long it has been under the threshold. A simple possibility, which also utilises the tangible channel, would be to place the object beside a screen that is able to retrieve and display this data.

Social Animalistic Objects do not only exhibit life-like reactions but emphasise the relation to the user through, for instance, adaption to the user beyond single interaction sessions or other continuity behaviour that may correspond to life-cycles of living entities, as for instance realised by the Furby doll. Other mechanisms could involve the Tamagotchi-effect, which motivates the user to turn attention to the device in order to sustain a required level of life parameters. Such an interaction paradigm can be useful in simple systems as stated above, but particularly in recurring scenarios of daily routines. This could be, for example, a plant container that expresses increasing contentment the longer the watering is kept on an ideal level. Or an alarm clock, which is satisfied when the inhabitant generally gets up quickly and in consequence uses pleasant and relaxed sounds as alarms, whereas annoyance due to ”pain” takes place when the user tends to hit the ”snooze” button often. Changing behaviour over time can therefore not only reflect changing internal states of an object, but also provide additional incentives for the user to act in certain ways, if it surprises, entertains or intrigues the user. This leads to the notion of *persuasive interfaces*, which has even more relevance on higher layers of anthropomorphism (see Section 4.3.3).

4.3.2 Talking Objects

The next level of anthropomorphic smart objects crosses the animalistic domain and employs interaction with more human-like objects through natural language dialogues. Although concepts of personality and also folk psychology can theoretically be employed without natural language, the speech modality provides rich means of realising anthropomorphism on these levels.

Emotional Talking Objects apply emotions and/or moods in a more holistic manner than is the case with animalistic objects, which already reflect a sense of emotion with the notion of affective behaviour, but only in terms of basic desires. The primary purpose of this is that emotion helps to increase the believability of anthropomorphic interfaces. Artificial emotion can also provide feedback to the user, such as indicating the internal states, goals and (to some extent) intentions.

The intention here is to support a mental model that results from folk-psychological processes of explaining *"how perceptions, beliefs, goals, intentions and actions relate"* [Persson et al., 2000]. Main element for such a model is the *Appraisal theory*, which formulates the idea that emotions are formed from our evaluations (appraisals) of events, which cause particular reactions in different people. The interdependencies between personality, moods and emotions are for instance modelled by ALMA [Gebhard, 2005], which was also applied in the Virtual Human project (see Section 3.3.1.1). Emotions can be displayed by different means, including facial expressions and body language of virtual characters (e.g. [André et al., 2000]) or also, more valuable for smart object scenarios, through speech as in [Schröder, 2001]. With such a module that describes and manages the effects of personality and emotion on a time axis, we will be able to further create talking objects with a believable model of their behaviour for short- and midterm interaction processes.

Combining speech and tangible interaction will facilitate more complex systems than the previous class, but also requires a more complex design and implementation process. Realising **Talking Objects with Personality** does not only support the anthropomorphic aspect, but holds additional benefits in speech-based interaction, if implemented with care:

Previous studies have shown that interacting with embodied conversational agents that have consistent personalities is not only more fun, but also lets users perceive such agents as more useful than agents without consistent personalities [Nass et al., 2000, Duggan and Deegan, 2003]. Furthermore is has been shown that the speech of a consistent personality enables the listener to memorise spoken contents easier and moreover reduces the overall cognitive load [Fiske and Taylor, 1991, Nass et al., 2000]. Similar findings have been made concerning human computer interaction employing speech synthesis: The computer was evaluated as more believable and the communication as more pleasant, whenever the synthesised voice characteristics fitted the speech output contents [Nass and Lee, 2001] and the perceived personality.

Personalities can be modelled by different means - for instance by sophisticated personality models as the five factor model [McRae and John, 1992] or simplified *traits* that summarise impressions about a person, such as shy, extrovert or aggressive. The chosen modelling method and level of detail depend on the application scenario and the intended effect on the user. One possibility is to align the personality to qualities of the object, e.g. a product in a shopping context, with its speech aligned to its brand image (see also Section 8.2). Another possibility is to model it after the user's personality, as suggested by the *Similarity-Attraction-Principle*, which claims that people prefer personalities that are similar to their own [Byrne and Griffitt, 1969]. Alternatively, theories from personality psychology propose the principle of *interpersonal complementarity* which specifies how a person's interpersonal behaviour evokes behaviour from an interactional partner [Kiesler, 1983]. In this notion, correspondence tends to occur on the so called *affiliation axis* (friendliness invites friendliness, and hostility invites hostility), while reciprocity tends to occur on the *power axis* (dominance invites submission and vice versa). A simple example can be found in the Luxalive lamp in Section 3.3.2.2, which builds on the latter phenomenon on the power axis.

Social stereotypes, which are generated by mix of assumptions towards people, are an option as well, as outlined in [Persson et al., 2000]. Typical cases are occupancy roles (doctors, waiters, police) or family roles schemas, all applicable to situations when the role is a dominant factor in the user-object relationship. Such stereotypes are another way to leverage expectations on goals, beliefs, morals, and behaviour of entities and might simplify the process of modelling distinct personalities.

This class of anthropomorphic smart object interaction makes sense for more complex functions that can be effectively interfaced through speech as the main modality. It still allows for intuitive, occasional interaction, like product inquiries in a shopping context (e.g. [Wasinger and Wahlster, 2005] [Schmitz et al., 2009]), exhibitions or providing instructions for programmable home appliances (e.g. video recorder).

4.3.3 Bonding Objects

The class of bonding objects builds on *Emotional Anthropomorphism* as described in Section 2.4.4 and basically comprises the social relation of the interactive device to the user. Talking Objects can generally be utilised as a base interaction paradigm in order to establish a relationship. The key aspect of this class is that relationship is a persistent and evolving construct that develops over time and spans multiple interaction sessions.

Hong et al. have summarised in their overview paper "Advances in Tangible Interaction and Ubiquitous Virtual Reality" that criticism within the TUI community was formulated towards the fact that research activities focus on short-duration application while it appears worthwhile to to consider longer periods of interaction in the user's life [Hong et al., 2008]. It has also been stated in this paper that *"tangibles could open new avenues for bonding with products, allowing products to physically evolve over time"*. This is further supported by Hornecker, who referred to studies showing that bodily interaction and touch rouse emotions and enhance inner sympathy [Hornecker and Bruns, 2004].

Only few studies on embodied agents have taken the endeavours to examine such long term effects on users. Bickmore and Picard have investigated long-term relationships in HCI within the FitTrack project (see Section 3.3.1.3) and recapitulated strategies for social relationship management - mainly for reducing social distance - that could be employed by a computer [Bickmore and Picard, 2005], which we have adapted for anthropomorphic smart object scenarios:

Social Dialogue: A form of talk that is particularly lacking in task-oriented content, also called *small talk* or *phatic communication*. Such a social dialogue can be used to maintain relational nuances even when no explicit task is being performed [Jakobson, 1960].

Social Deixis: The encoding of relational status in language is a phenomenon called *social deixis* [Levinson, 1983]. A familiar example is the form of address, greeting and parting routines that are used between people having different relationships, with titles ranging from professional forms (”Dr. Smith”) to first names (”Joe”) and greetings ranging from a simple ”Hello” to the more formal ”Good Morning”. A further example is *politeness theory*, which describes different forms of indirectness for a request depending on the nature of the relationship between the requesting person and potential supporter and on how cumbersome the request is (e.g. the difference between asking your boss for $5 or a close friend). Thus, appropriate use of social deixis can help to affirm and maintain the status of an existing relationship, while using language features indicating a different form of relationship can signal a desire to change the relational status.

Empathy: Delineating the process of attending to, understanding, and responding to another person’s expressions of emotion is not only important for intimate relationships but also cited as most important for building good working alliances between helpers and clients or physicians and patients. Klein et al. have shown that an empathetic and accurate computer can achieve significant behavioural effects on a user, similar to what could be expected from human empathy [Klein et al., 2002].

Self-Disclosure: Reciprocal self-disclosure increases trust, closeness and liking, which has been demonstrated in text-based human-computer interaction. Users are also more likely to buy a product from the computer if reciprocal, deepening self-disclosure is applied [Moon, 2000].

Continuity behaviour: Bridging the time people are apart can be facilitated by appropriate greetings and farewells and talk about the time in separation. Such behaviour is important to maintain a feeling of persistence in a relationship [Gilbertson et al., 1998].

Humour: Computer systems that employ humour are rated as more likeable, competent and cooperative and thus form an important relationship management strategy (e.g. [Morkes et al., 1999]).

Reeves and Nass have further demonstrated several relational effects, which result in a more sympathetic attitude towards the computer if it [Reeves and Nass, 1996]...

- ...uses flattery, or praises rather than criticises users.
- ...praises other computers or criticises itself, instead of praising itself or criticising others.
- ...becomes more like the user over time or if it maintains a consistent level of similarity.
- ...*teams* with a user, which can be achieved by simply signifying that user and computer are a team - which will also increases the willingness for cooperation.

All of these strategies can be employed by conversational means (i.e. Talking Objects), while some can also be implemented in a more subtle manner by animalistic processes, e.g. expressing empathy towards the user. The application of such methods does at the same time address joy of use qualities that relate to the individuality of the interface, namely the key factors *personal experiences of use* and *individualisation*, and furthermore the desire for social interaction (see Section 2.2.1.1).

The domain of applications for bonding objects can be derived from the range of scenarios that relationships are generally good for, as provided by models of social psychology. These include emotional support (e.g., esteem, reassurance of worth, affection, attachment, intimacy), appraisal support (e.g., advice and guidance, information, feedback), instrumental support (e.g., material assistance), group belonging, opportunities to nurture, autonomy support, and social network support (e.g., providing introductions to other people) [Berscheid and Reis, 1998].

In particular, relationships can also play a role in **persuasive scenarios**. Perceived trustworthiness and overall charm of a source of persuasive information play an important role in the *Elaboration Likelihood Model of persuasion* [Petty and Wegener, 1998]. In this theory, if a decision is of low personal importance, then characteristics such as trustability and sympathy of the source of information have significant influence on the decision. Thus, bonding objects could be used in sales scenarios, which aim at building relationships with client users just as one could expect from a human sales person [Anselmi and Zemanek, 1997].

Education scenarios are another possible field of application. It has been shown that within elementary school, students feelings of relatedness towards both their teacher and classmates strongly influence their cognitive, behavioural, and emotional engagement in classroom activities [Stipek, 1996]. Furthermore, there is evidence that relationships between students are important in peer learning situations, including peer tutoring and peer collaborative learning methodologies [Damon and Phelps, 1989].

In studies of **customer service relations**, there is differentiation between three types of relationships: *Service relationships*, in which customer and service provider expect to meet again in future. *Pseudorelationships*, in which customers expect to interact again with the same company but not the same person in the future and *service encounters*, which do not presume expectations of future interactions. Gutek et al. [Gutek et al., 2000] observed in their studies that customers in a service relationship reported more trust in their service providers, more interest in continuing the interaction, and more willingness to refer the provider to others, than customers in either pseudorelationships or service encounters. Results also indicate that a service relationship is significantly more effective at inducing trust, commitment and referrals than attempts to establish brand or firm loyalty.

Summing up, qualities of relationship can have major impact on tasks in diverse areas, such as education, sales, or different types of service situations. Thus, employing bonding objects in such and similar contexts does not only have the potential to create more believable and enjoyable smart object interaction, but also for significantly improving user performance and results.

4.4 Anthropomorphic Interpretation of Tangible Input

Employing an anthropomorphic stance towards a smart object opens a new metaphorical space for tangible interaction that provides a high intuitiveness and a natural understanding due to the inherent attributes of anthropomorphism. Traditional types of tangible input typically consist of specialised gestures and movements that relate to the particular application and its functions. On an anthropomorphic level, tangible input can follow an easily interpreted and embracing metaphor, which obtains meaning by the physiological and emotional impact of the motion. In Table 4.1 we provide a compilation of tangible input types and suggest interpretations for each in the anthropomorphic domain, based on typical physiological and emotional reactions / intentions derived from [MacLean, 2000b]. Furthermore we propose functions for each mapping in order to inspire interaction designers to develop anthropomorphic interaction through tangible interaction. We use cube affordances determined through user observations by Sheridan et al. [Sheridan et al., 2003] as a starting point for a base set of interactions and extended it with actions particularly applicable in an anthropomorphic context.

Tangible Input	Anthropomorphic Interpretation	Functions / Effects
Touch or Presence	Pay attention to object, calm down	Activate, provide further data
Pick up	Pay attention, connect, observe	Provide further data, stop current action
Point at/Orient/Place near object	show something to object	perform action related to shown object (e.g. connect, interact or comment)
Shake	Wake up, punish	Request, initiate, reset, stop current action, undo last action
Throw away/Drop	Release, send away, abandon	stop functionality, reset, undo last action, mute object, turn off
Tap/Knock	Request attention, wake up, signal	Trigger action, request information, repeat last action
Rub/Pet	Calm down, reward	Notify about correction of bad state, acknowledge useful information or action
Hit	Punish	Disapprove last action, reset
Squeeze/Pinch	Engage, force, express anger, threaten	Activate, trigger action, mute, turn off
Spin/Rotate	Confuse, irritate	Shuffle data, trigger random action
Cover/shade with hands	Protect, Hide	Recover from erroneous state, shut down

***Table 4.1**: Anthropomorphic interpretation of tangible input*

The listed functionalities for each input type can only be considered as general suggestions, while the anthropomorphic interpretation provides a notion of life-likeliness that can be exploited and integrated into the overall interaction design. Certainly, the listed input methods do not have to be implemented as anthropomorphic stimuli, instead they can be replaced by proprietary functional effects or mixed with such. For example, an anthropomorphic USB memory stick might encrypt its contents when squeezed (functional interpretation), confirming the process with sounds that express exertion and effort (anthropomorphic reaction) and finally confirm the successful encryption with joyful sound (anthropomorphic and functional). Moreover, certain types of behaviour might result in a variation of the proposed interpretations, in particular for those actions that have a rather neutral connotation. For example, tapping could either be interpreted as a request for attention but also cause fear in an extremely shy object.

For the realisation of anthropomorphic tangible input we compiled a list of off-the-shelf-sensors that can be integrated into smart objects and analysed their applicability for tangible input recognition, which is presented in Table 4.2. We further indicated how important the spatial arrangement within the object is, such that it has to be considered in the physical design process to facilitate the intended functionality.

Tangible Input	Sensor Solution	Importance of Placement
Touch	Capacitive/Resistive touch sensor	High
	Pressure sensitive conductive rubber sheets	High
User presence	Infrared motion/distance sensor	High
	Ultrasonic range finder	High
	Infrared Thermometer	High
Pick up	Acceleration sensor	Medium
	RFID	High
	Ball/Binary movement sensor	Low
Point at / Place near object	Combining ultrasonic/infrared	
	sender and receiver	High
	external or internal cameras	High
Shake	Acceleration sensor	Medium
	Ball/Binary movement sensor	Low
Throw away / Drop	Acceleration sensor	Medium
Tap/Knock	Piezo vibration sensor	Low
	Electret microphone	Medium
Rub/Pet	Capacitive touch slider	High
Hit	Electret microphone	Medium
	Acceleration sensor	Medium
Squeeze/Push down	Flex sensor	High
	Pressure sensor	High
Spin/Rotate	Acceleration sensor	Medium
Cover/shade with hands	Light sensor/Photoresistor	High
	see also *User presence*	

***Table 4.2**: Sensor solutions for tangible input recognition*

4.5 Synopsis

The main goal of this chapter was to structure and conceptualise the primary design dimensions relevant for the development process of tangible, anthropomorphic smart objects and to develop guidelines for interaction designers. For this purpose we have analysed research and case studies in respective fields and discussed the rationale and justification for the formulated design guidelines.

The general design rules are primarily important to decide whether the concept of anthropomorphic smart objects is appropriate at all for a given application. This also involves building up a consistent mental model, which is a fundamental process of understanding the interaction and thus a main requirement for successful human-computer interaction. We have also discussed various task domains and their applicability for the proposed interaction paradigm. In conclusion, we excluded certain application types, such as safety-critical applications in professional work environments and outlined the potential benefits for use cases that suit well to anthropomorphic smart objects. There are several aspects regarding the traditional use of an object that have to be taken into account, when new functionality and interactivity is added. *Function*, *action* and *modality* are design dimensions where conflicts with the original use of the augmented object may arise. In each of these three dimensions the relationship between traditional and extended attributes plays a significant role in the overall design process, which makes it important for the interaction designer to understand and consider the described phenomena.

We have seen that the attribution of anthropomorphism to an artefact can be achieved by several means, such as adjusting the visual appearance, using a speech interface or life-like sounds and haptics, or employing pro-active behaviour, which can moreover be modelled in stereotypical ways that allow to recognise common behavioural patterns, which are known from everyday encounters with other people. Industrial designers can make essential contributions to strengthen the anthropomorphic stance, by matching the appearance of the object to both the anthropomorphic paradigm and possibilities of tangible input. Most of all, the notion of anthropomorphic affordances constitutes a novel concept that extends the traditional understanding of affordances by the perceived possibilities of use that result from anthropomorphically designed elements.

Technological components, which implement interaction recognition, application logic or output generation, can be integrated into the object itself or outsourced into the environment. We defined four discrete classes of different levels of object instrumentation and discussed the advantages and drawbacks of these approaches.

To support developers and designers of smart object systems to decide which type of anthropomorphism is adequate for the intended application scenario, we have constructed the first classification of anthropomorphic smart objects with their distinct properties, structuring existing and future design concepts. These properties include typical application scenarios as well as strategies of designing interaction concepts for each class.

The diagram displayed in Figure 4.4 summarises this classification. The graph reflects the *anthropomorphic key attributes* of a *class* and their interdependencies, whereas each of these attributes has certain advantages and features, which were introduced in Section 4.3. These beneficial properties support certain fields of application or scenarios. To reduce the complexity of this visual representation, it only connects anthropomorphic key attributes to a certain advantage or feature with a directed arrow, if this advantage is distinct for this attribute and not yet facilitated by another attribute of a less complex object class. Likewise, arrows are only drawn from advantages to fields of application, which are not yet supported by features by another class further to the left.

Furthermore, a new dimension to tangible input design emerges through the anthropomorphic approach. We elaborated tangible input elements and proposed a mapping to an anthropomorphic interpretation and appropriate functions on a system level to inspire the creation of engaging and meaningful interaction concepts. In addition to that we compiled a list of sensor technologies and their capability for user input recognition when integrated into smart objects.

The diagram in Figure 4.5 shows a decision tree that summarises the main design choices introduced in this chapter.

We will walk through this decision tree by means of an example, the Digital Somme-

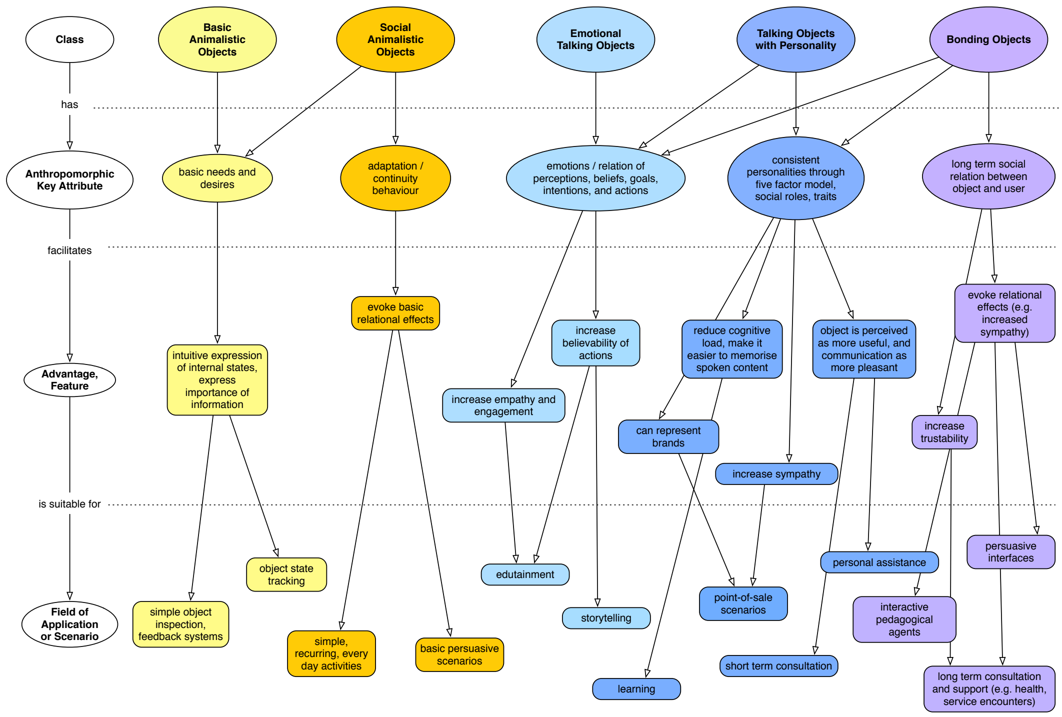


***Figure 4.4**: Anthropomorphic smart object classes, key attributes and fields of application*

lier (DigiSom), which we have developed for exhibitions and retail environments (more in Section 8.2): The first question to be answered in the planning process is whether the system employs a single object that acts as an independent interaction partner. The DigiSom involves multiple bottles that will be extended to smart objects, hence, this question has to be answered with ”no”, such that we have to consider whether a mental model is still appropriate in this scenario (see Section 4.1.1). We intended to have one-on-one interaction situations, in which each bottle interacts independently with the user. Thus, the user does not interact with distributed components but focuses on a single object at a time. The next step involves the question, whether the task involves expert users (see Section 4.1.2). The DigiSom is situated in a retail environment, where we expect casual interaction of novice users - basically customers that interact once or twice with the system in total. Then we have to decide, which type of interaction we plan to employ in the system. Since we do not want autonomous behaviour, we assume that some sort of dialogic interaction is most appropriate and therefore decide to employ talking objects with personality. The encounters of customers with the application are expected to be of a rather short duration, for which reason we do not integrate more complex emotional models or bonding strategies in this use case scenario (see also Section 4.3).

We believe that the overall concept has the inherent potential for simple and inspiring interaction that provokes a high degree of curiosity. The guidelines developed in this chap-

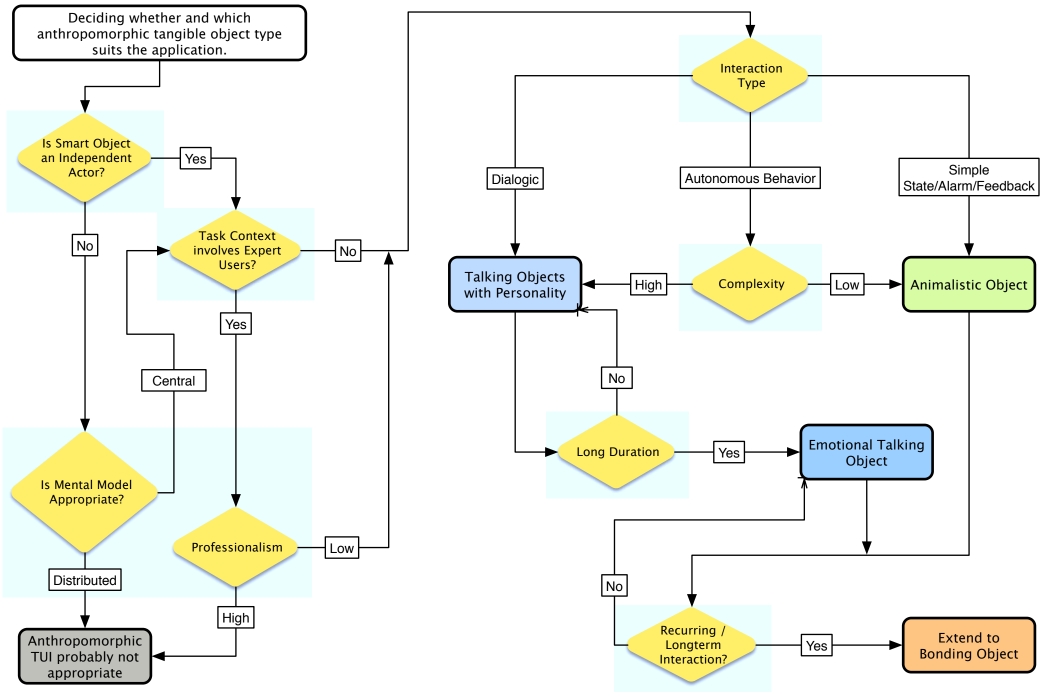


Figure 4.5: *Decision tree for deciding which approach suits the given task. The colours are chosen in accordance to Figure 4.3.*

ter also regard the quality criteria of joy of use, introduced in Section 2.2.1.1 and therefore address essential concepts for developing engaging, emotional and joyful interaction. Also, quality criteria for TUI stated by Hornecker [Hornecker, 2004] are addressed as well in this chapter, particularly by analysing several aspects of affordances, while at the same time adhering to Fishkin's taxonomy.

As we have seen, the technical implementation of interactive smart object systems can take place in many variations of embedment and distribution. In the next chapter we will introduce the architecture of our testbed for realising TASOs in instrumented environments, addressing the needs of all four levels of distributing technology and of all proposed types of anthropomorphism.

5 PROTOTYPING FRAMEWORK

In this chapter we will describe the structure and implementation of the *TASO Platform*, our testbed for the development of TASOs as described in the previous chapter. One major goal in the development process was to create a prototyping framework that allows to easily utilise a combination of technologies embedded in objects and in the environment. It is furthermore crucial to be able to exchange technologies, such as sensor platforms with minimal changes in reused modules. We also aimed at a level of modularity that allows a high degree of reuse of software components for future setups of new smart object applications.

In Section 5.1 we will start with introducing sensor platforms and describe the sensor processing framework *TASO SensPro* with the corresponding graphical user interface, which constitute the main part the sensor-based prototyping environment [Schmitz et al., 2008a]. This is followed by an overview of the overall architecture in Section 5.2 of our prototyping environment and continued with more detailed descriptions of its major components. Furthermore, in Section 5.3 we present services of the instrumented environment that support the development of user interfaces, mainly the camera-based user tracker *MaMUT* (Section 5.3.1.1), the *Presentation Manager* (Section 5.3.2) and the spatial audio system *SAFIR* (Section 5.3.3). We will conclude this chapter with a summary in Section 5.4.

5.1 TASO SensPro - Sensor-Based Prototyping of Smart Object Applications

Embedding sensing, computing and communication capabilities into physical artefacts is the essential building block for the creation of smart objects. Emerging sensor technologies that go beyond RFID recognition are becoming cheaper and smaller, paving their way into objects of everyday life. The potential ubiquity of wireless sensors will particularly affect future generations of products as elements of production, processing and delivery chains, enabling them to sense and record their state and communicate with their environment. The prospective immediate benefits for manufacture and retail are an essential factor that is increasing the chances for research work in this field to get adopted by industry - just as RFID technology has already made the step into the commercial sector, as it provides economic values by facilitating process automatisation such as product tracking or inventory stocktaking.

Research in the field of wireless sensor networks has brought up commercial off-the-shelf systems, which provide sensor nodes with integrated sensing, computing and communication capabilities. Some of them can be configured only by means of basic parameters, others have a programmable microcontroller and can be flashed with custom firmware. Such devices use their computing resources for sophisticated tasks like preprocessing sensor readings or routing of messages through the network. Another emerging research trend involves smart materials or also called sensorial materials, which are conceived as materials containing spatially distributed sensor elements. Produced as precursor materials, such components can be processed as integral parts of future products, as opposed to a subsequent instrumentation with sensor modules. Advancements in this research area will facilitate additional means of pervasively embedding sensors into our physical environment and further support the vision of ubiquitous smart objects (see e.g. [Delaney and Dobson, 2009]).

5.1.1 Sensor Platforms

Wireless sensor actor networks (WSANs) in general consist of a set of sensors and actuators that are wirelessly connected and are employed in a variety of application domains, ranging from observing the habitat of animals and smart home systems to military applications that involve observation of battle grounds and combat activities. General requirements of WSANs include fault tolerance, quality of service (e.g. drop-rate of network packages, throughput), robustness (against e.g. failure of single nodes), scalability, dynamic network-reconfiguration and programmability. Some of these requirements are less important in a prototyping context as described in this chapter. For instance, scalability will normally not be relevant, since we focus on systems that do not involve more than a couple of sensor nodes. There are sensor platforms such as the Arduino[1] system that do not regard many of the aforementioned aspects regarding networking and other aspects of large scale field use. Instead, they offer a low-cost hardware platform for prototyping custom applications on a microcontroller basis.

[1] `http://www.arduino.cc/`, last visited June 30, 2010

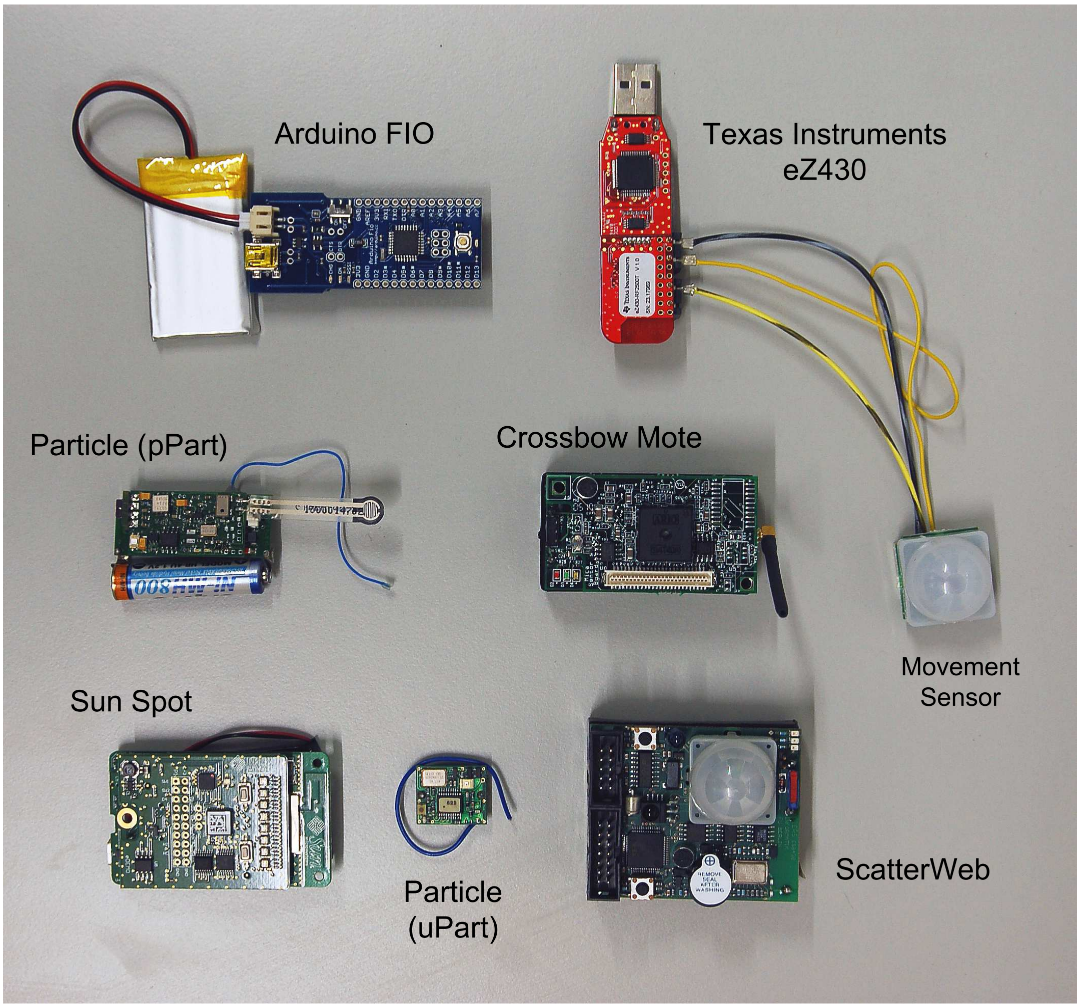


Figure 5.1: *Sensor nodes experimented with during the development of TASO SensPro*

A sensor node consists of several basic components: A **central processing unit** (CPU) to process data and to execute the actual sensor application, **memory** to store the application and data (often realised with different memory types for each type of data), **sensors/actors** that collect environmental data, **communication interface** for wireless communication with other sensor nodes or a base station, and **energy supply**. No single sensor platform can meet all the requirements of different applications. Therefore we experimented with various wireless sensor platforms, including the μPart, the MICAz, the aforementioned Arduino nodes, and ScatterWeb, while using the latter system only for environmental observations, such as measuring the temperature in a room or detecting the presence of people (see Figure 5.1 to see a collection of different sensor platforms). The μPart platform has been initially developed at Karlsruhe University [Beigl et al., 2005]. It is produced and sold by the spin off Particle Computer GmbH. μPart nodes provide sensors for light, movement and temperature. They can be configured to set the sampling rate and other basic parameters through light sensors, but firmware modifications are not possible this way, preloaded software merely sends out data via a proprietary radio communication protocol, which can be received by compatible bridges. The MICAz platform is a Zigbee compliant implementation of the Mica Mote platform [Hill and Culler, 2002], developed at Berkeley University in cooperation with Intel. MICAz nodes are sold by Crossbow Inc. and are shipped without integrated sensors, but they can be stacked with various sensor boards. We use them mainly in connection with the MTS300 board, which provides sensors for light, temperature and noise as well as a dual-axis accelerometer and a dual-axis magnetometer. The ATmega128L microcontroller on the nodes provides computing resources to run custom firmware from the integrated 128 KBytes of flash memory. Table 5.1.1 provides an overview of all sensor platforms, which we have experimented with during the course of this work. The implementation of the sensor integration described in the following is based on a framework that we have designed for rapid development of sensor-based applications.

Name	Onboard Sensors	Onboard Actuators	Network Topology	CPU	RAM/Flash	Connectors	Development / API
Texas Instruments eZ430-RF2500	Temp	2 LEDs	Star	MSP430F2274 at 16MHz	1KB/32KB	through additional sensor board	direct CPU-Programming
Crossbow Motes (MICAz XM2110+MTS420)	2-axis acc., light, temp/humdity, pressure, GPS	-	Mesh, other	ATMega 1281 at 16MHz	8kb/512kb	I/O board required (MDA 300, incl. temp./humidity)	nesC/TinyOS (Java-Interface)
Particle Computer (pParts)	light / IR light, temp, sound intensity, 2/3-axis acc.	2 LEDs, buzzer	Mesh	PIC 18F6720 at 20 MHz	128kb program Flash, 512 kb FLASH for data	no	Java library, C (onboard)
Particle Computer (uParts)	movement, light, temp	1 LED	Mesh	12F675 at 4 MHz	Flash 1.4 kByte, SRAM 64byte, EEPROM 128 Byte	no	Java library, C (onboard)
Scatterweb	temp, motion (IR), vibration, light	1 LED	Mesh	MSP430F1612IPM at 8 MHz	512 kByte EEPROM	13 digital I/O, 5 analog in, 1 analog out	Scatterweb SDK, optional: nesC/TinyOS
Sun Spots	3-axis acc , temp, light, 2 momentary switches	8 tri-color LEDs	Mesh, other	32-bit ARM920T at 180MHz	512K RAM 4MB Flash	5 I/O pins, 4 high current out	Java
Arduino FIO	none	-	Mesh (with XBee)	ATmega328P at 8MHz	32 kb Flash	14 digital I/O	Arduino IDE
Arduino Pro Mini	none	none	Mesh (with XBee)	ATmega168, at 8Mhz	14 kb Flash	14 digital I/O, 6 can be used as analog	Arduino IDE

Table 5.1: *Overview of sensor platforms we experimented with*

5.1.2 Requirements for a Sensor-Based Prototyping Framework

So far applications have to access sensor data directly through hardware-dependent libraries. In many cases hardware details influence the entire software design, which complicates software maintenance especially when hardware is exchanged or updated. This particularly poses a burden in research environments, where different device platforms are interchanged and used for different applications, reusability of software modules is very restricted.

To enhance rapid development of sensor based applications, a solution had to be found, which provides a structure for data processing through reusable software components. From previous experience with sensor applications we have extracted the following requirements for such a framework:

1. Sensor data should be accessible through a high-level API instead of raw byte access.
2. Data processing should be as simple and efficient as possible to cope with high data rates from a large number of sensors simultaneously.
3. All sensor specific hardware and software details should be hidden.
4. A large number of data sources should be supported, including not only wireless sensor networks but also other types of input (e.g. reading from files, databases or other network sources).
5. There should be prebuilt components for common application tasks as well as for development related issues (e.g. data inspection, debugging).
6. The implementation should be open for extension, allowing further development and adaption of new sensor hardware.

5.1.3 Processing Sensor Data Streams

The design of the framework is based on the concept of stream processing (see [Stephens, 1997] for an overview). It uses a data driven approach where information is contained in possibly infinite streams of data entities. Each entity can be seen as an event and in our terms, each raw sensor reading is such an event. In our framework, events are produced, consumed and processed by three kinds of modules: *Input modules* encapsulate data sources such as sensor networks or files, and provide a steady flow of information into the framework. *Output modules* are for example used for displaying data, connecting to other applications or triggering actions based on events. In-between input and output optional *filtering modules* process the data streams to extract and prepare relevant and meaningful information from incoming sensor data. The data flow is implicitly uni-directional - if for any reason a bi-directional flow is required, it must be explicitly constructed. Figure 5.2 displays the encapsulation and abstraction steps of the sensor stream processing in TASO SensPro.

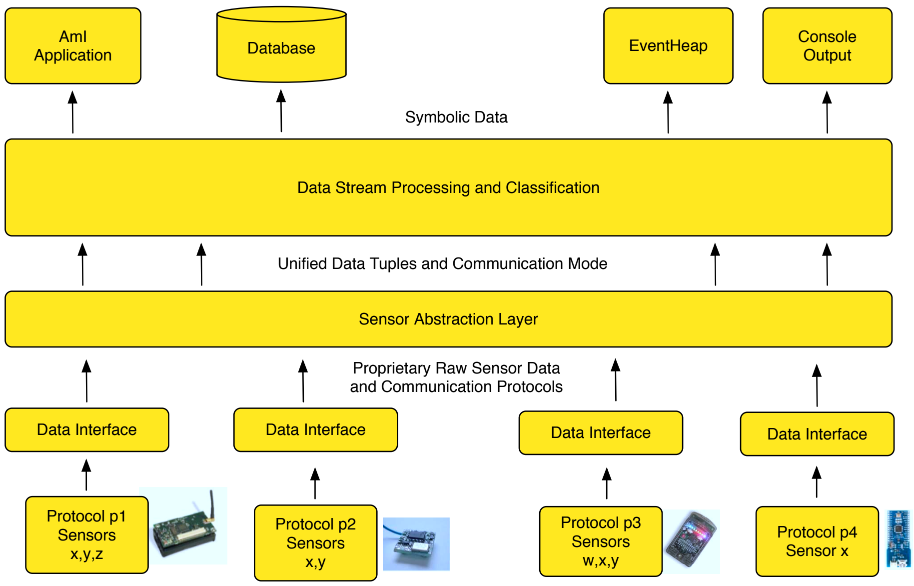


***Figure 5.2**: The abstraction layers of sensor stream processing in TASO SensPro*

The sensor-based component of a new application is built by choosing one or more modules of a set of modules provided by the framework and connecting their in- and outputs appropriately to construct a desired function. Although new custom modules can be easily implemented and integrated, the set of prebuilt modules covers many common data processing tasks, such that a wide range of simple applications can be developed quickly with the given components. As sensor nodes define their own message format, the encapsulating input modules are customised to suit the underlying sensor platform, and have to be extended, whenever new platforms have to be integrated. Such input modules provide flexibility and connectivity towards the framework.

There are already several tools available to support event processing for intelligent environments. As we will see in Section 5.2 we decided to use the EventHeap to share data between components of the infrastructure, but sending each sensor reading directly to the EventHeap is in most cases not an appropriate procedure, since a single reading does not contain information of high value. The raw sensor readings have to be preprocessed and analysed in order to generate higher level events that are meaningful enough to be sent to the actual application. Such events include user actions, which could for instance be accumulated through this interface to infer user *activities*.

5.1.4 Implementation

We decided to implement the framework in the Java programming language mainly because of two reasons: Most sensor networks already come with some Java library or classes giving a good starting point. Furthermore, Java is platform independent, which is particularly important when working in a heterogenous environment.

As already stated, there are three types of modules, namely *Input*-, *Output*- and *Filter*-Modules. They are linked together using a design pattern from object oriented programming called the *Listener Pattern*. It allows coupling of software components and uni-directional transmission of data from a subject (input) to one or more listeners (output).

This approach allows to separate the application architecture from the specific implementation of modules. So far, there are 56 prebuilt modules. Input can be obtained from different sensor networks, from files or from network or serial connections. Amongst others, output can be directly fused into files, databases or higher level systems like the EventHeap. For data inspection and debugging there are modules that display data either graphical of textual. Filter modules implement various computing functions as for instance sum, variance or standard deviation of a sequence of values. Other filters also help to organise the flow of sensor data, such as for instance drop repeating values, split message streams by origin or type, or translate certain field values of the messages. The translation for example is often used to map sensor IDs to products and can be configured through an external XML file, eliminating the need for touching the binaries when changing product mappings. Modules are implemented by extending abstract base classes, which already contain functionality for establishing links between modules, so there is minimal effort needed.

5.1.5 Special Purpose Modules

Not all algorithmic problems can be broken down into simple solutions. In some cases it is more reasonable to implement a complex function as a self contained module, which provides the necessary interfaces for connecting it to the framework. **Robust motion analysis** based on acceleration sensors is such a task, requiring sophisticated methods for information extraction. The complexity of the given data in this case requires a tailored approach, for which we generated a specialised module for our framework that employs machine learning algorithms of the WEKA toolkit [Witten and Frank, 2005]. The following example outlines the usage of this module on data generated by acceleration sensors preprocessed on the sensor board.

Acceleration data is sampled in a relatively high frequency e.g. with a rate of 100 samples per second, as on the MICAz nodes used in the DigiSom installation described in Section 8.2. In a first step the readings are preprocessed by computing several statistical values over a window of 30 samples corresponding to a time span of 300 ms. The statistical values are mean, variance, standard deviation as well as the number of local minima and maxima for each acceleration axis, forming a block of 10 characteristic values, which describe the state of acceleration for a specific time interval. These blocks are computed by our own firmware installed on the sensor boards, and are sent to a PC that runs the sensor framework and then processed by the integrated WEKA library - first in a training phase and after that in a classification phase. In the training phase, several classes of movement states have to be specified, and incoming readings are assigned to the according movement class. Based on this training data the system can classify incoming data packages and send results further into the framework. Using a statistical approach of movement state classification, we are able to differentiate certain movements regarding their intensity (no movement, slow movement, fast movement), their direction or anything else that can be extracted from the set of statistical values. The classification is done by the WEKA implementation of a Ripple-Down Rules (RDR) classifier [Comption and Jansen, 1990], but other classifiers like C4.5 decision trees [Quinlan, 1993] have also proven to be useful. For a user of our framework, there is no need to write a single line of code during this process. The library contains specialised modules that allow a quick integration of WEKA classifiers into the given framework.

5.1.6 Visual Prototyping

A graphical user interface has been developed to further accelerate the prototyping process, particularly when microcontroller-based sensor systems are involved. One of the goals in the development was to enable developers to quickly setup and assess sensor systems with minimal efforts. The GUI provides the means to build up stream-based sensor interpretation basically by dragging graphical nodes onto the working canvas (which corresponds to the instantiation of filter classes) and connecting nodes with directed arrows (which corresponds to connecting input and output of filter classes) in order to direct the flow of data messages within the sensor interpretation stream. Figure 5.3 shows an example setup of the GUI in action.

The class attributes of filters are set through text input fields. Special nodes that support the aggregation of learning data for WEKA-based machine learning processes are introduced and help to create data in the WEKA compliant *Attribute-Relation File Format* (ARFF).

The features of the graphical user interface are briefly summarised in the following:

Click'n Play: When the system has been started after setting up nodes, the GUI starts all sensor data streams and displays intermediate values and results instantaneously. Adjusting parameters of active nodes can be done during runtime. Thus, the impacts of changes in the sensor processing stream are immediately displayed throughout the network of nodes, such that tests and debugging can be conveniently and efficiently performed.

Code export: Each instance of a combination of nodes processing sensor data created with the GUI can be used as a standalone application that forwards results of the sensor processing layer to other software components such as databases, text files or tuple spaces (e.g. the EventHeap). Furthermore, the resulting network of nodes can be exported as packages that can either be integrated into other Java applications or also started as a standalone application on other machines.

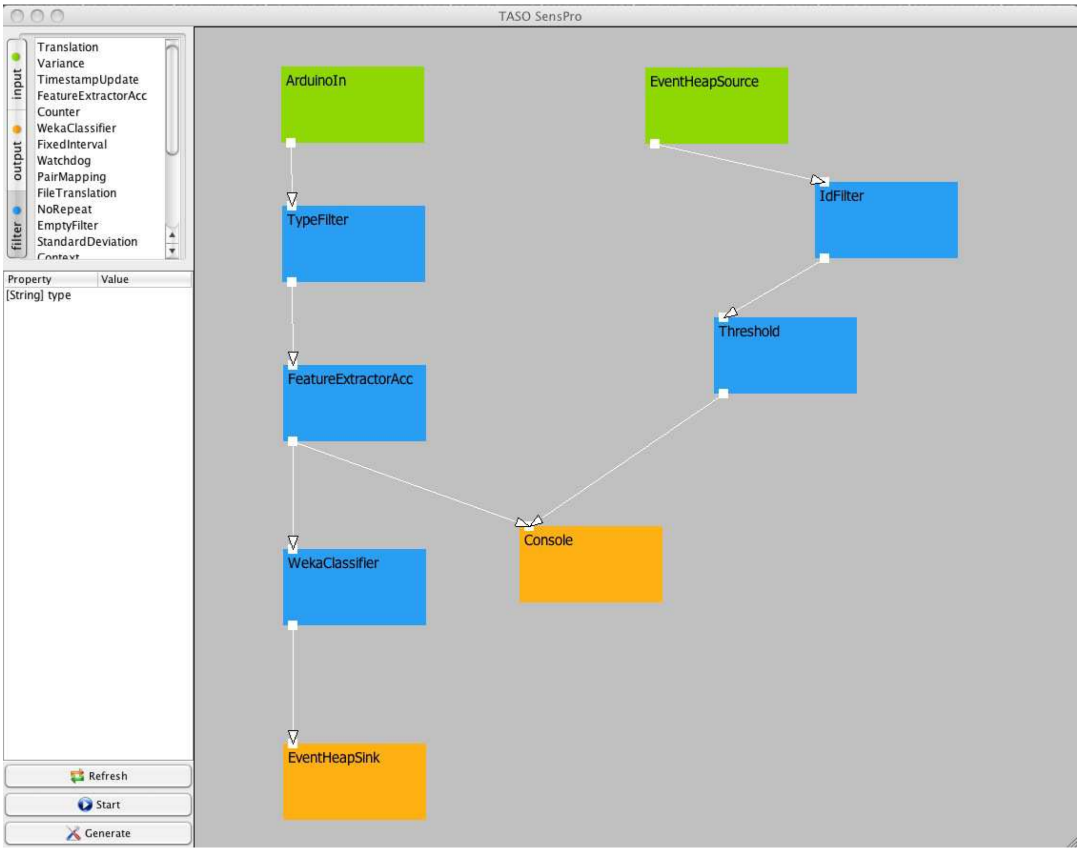


***Figure 5.3**: The graphical user interface of TASO SensPro*

Automatic integration of new filters: When new code is added to the sensor framework in the form of new filters that e.g implement singular processing steps, it is automatically detected by the GUI and made available to the user if it is added to the Java packages of the sensor framework or when the new class inherits from key (abstract) classes or implements existing interfaces.

Multi-threading: The ability to fuse and split data streams is done through a multi-threading approach and synchronisation of threads through atomic access on data structures.

5.2 Overall Architecture of the TASO Platform

We have started the development of the TASO Platform in the **S**aarland **U**niversity **P**ervasive **I**nstrumented **E**nvironment (SUPIE) as described in [Butz and Krüger, 2006]. It consists of several general purpose in- and output devices, which are shared by several applications. The software architecture of SUPIE has originally been designed for the seamless integration of various services and applications supporting different tasks such as shopping assistance or indoor navigation systems. Some elements of SUPIE are reused and extended for the TASO Platform.

The latest developments are also integrated into the Innovative Retail Lab (IRL) [Krüger et al., 2010], an application-oriented research laboratory of the DFKI run in collaboration with the German retailer GLOBUS SB-Warenhaus Holding, focusing on intelligent shopping assistance (see Figure 5.4).

Figure 5.4: *The instrumented environment of the IRL*

Figure 5.5 provides an abstract overview of the logical system layers, each of which informed by the concepts developed in Chapter 4. The lowest layer manages and processes the instrumentation of the environment and smart objects, regarding the guidelines for instrumentation and sensor data interpretation. The abstraction of hardware-dependent data generated by smart objects as described in Section 5.1 is the main task of this module. Services of the

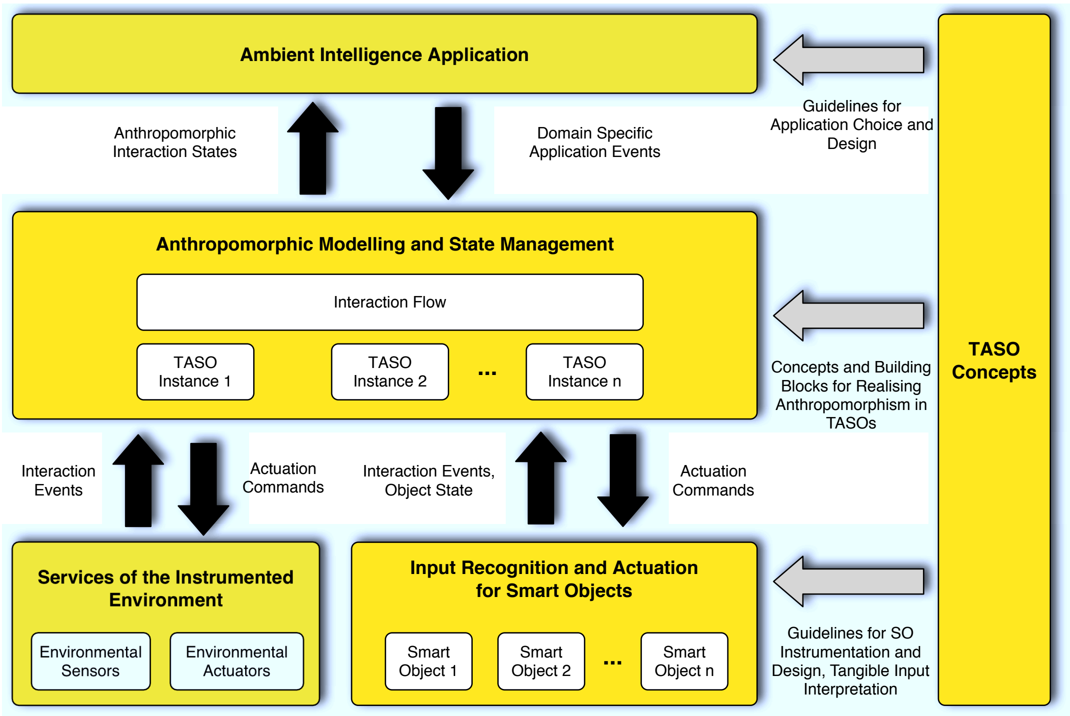


***Figure 5.5**: Abstract overview of the prototyping framework*

instrumented environment are realised on this layer as well, providing access to additional sensors (e.g. cameras) and actuators (e.g. visual and auditory displays). The layer on top regards the anthropomorphisation of smart objects and creates TASO instances, building on the conceptual results previously introduced. This also integrates the expression of personality and/or emotions, for which we will introduce novel means in Chapter 6. The overall flow of interaction is controlled here as well, implementing anthropomorphic behaviour on the basis of high-level user interaction events provided by the instrumented environment and smart objects. The overall AmI application layer adds domain specific components to complete the end-user application, based on the anthropomorphic interaction modelled on the layer below. The modular overall organisation principally supports the rapid development of AmI multi-user applications [Schmitz, 2006] [Stahl et al., 2005]. The architecture particularly allows the quick integration of new generations of hardware and software components.

The system architecture of the resulting TASO Platform is displayed in Figure 5.6 and described in the following. Yellow boxes comprise elements that have been developed during the course of this dissertation, while blue boxes represent re-used systems, which have been integrated into the prototyping architecture.

TASO SensPro is the major component of this architecture and deals with the integration of smart object sensor systems, which was explained in more detail in Section 5.1. It basically extracts more abstract information from sensor readings of instrumented objects

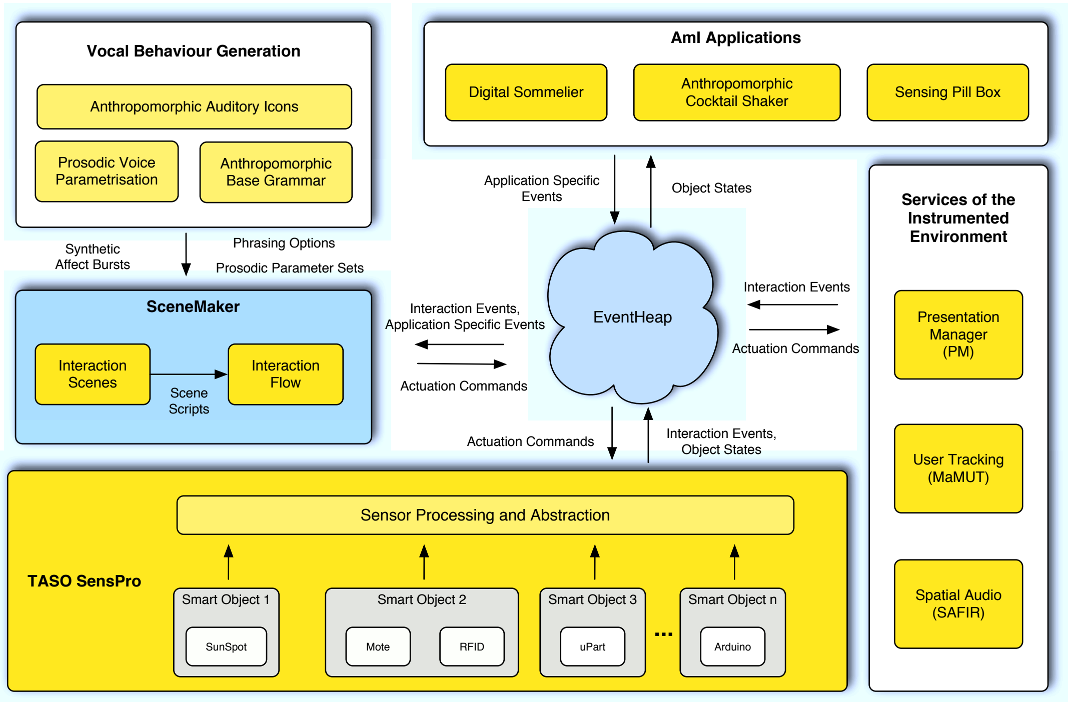


***Figure 5.6**: The architecture of the TASO Platform*

and delivers this information to the EventHeap, where it is generally available for all running applications. The sensor processing can take place on the sensor board itself or also on one or more *bridges*, which are computers that receive data from sensor boards to process and feed them into the architecture. The sensor-processing modules can be composed from pre-programmed components (the sensor filter library) and even a graphical user interface to reduce prototyping efforts. Actuation commands will be received and processed to trigger output rendering on the actual device, such as sounds or vibration. The setup of the architecture allows to run several instances of sensor-processing on different hosts, since all processed data can be communicated over the shared data space.

The **SceneMaker** module deals with the management of internal states of anthropomorphic entities and the overall interaction flow [Gebhard et al., 2003]. The SceneMaker toolkit enables non-expert users to compose interactive performances of virtual characters in a rapid prototyping approach in two steps: A scene flow is defined with finite state machines in a first step and in a second step, the contents of each scene are developed, which is typically done with a simple scripting language. Although this toolkit is designed for interactive performances of animated agents, it can also be used to model the behaviour of other entities in a finite-state-machine-based manner, such as anthropomorphic smart objects. For the integration of the SceneMaker we have developed extensions that basically facilitate the communication between the SceneMaker state machine and the remaining components of our architecture. The SceneMaker mainly receives interaction events released by TASO SensPro and the services of the environment to control the interaction flow and triggers output generation as pre-defined in the according interaction scene script.

The **Vocal Behaviour Component** provides building blocks for the generation of anthropomorphic expressions, including prosodic parametrisation (Section 6.2) of speech output for personality expression, an anthropomorphic base grammar (Section 6.3), and anthropomorphic auditory icons (Section 6.1). These elements are employed by interaction scenes defined by the SceneMaker.

Services of the Instrumented Environment provide services of different kinds for multiple concurrent applications at the same time. This includes for instance the user tracking service MaMUT that determines the location of users on different resolution levels (see Section 5.3.1.1) or the Presentation Manager (PM), which manages and allocates display resources to applications. This separation of services from applications also hides the technical complexity of such services behind a simple interface. For example, a shopping assistant such as the Digital Sommelier (see Section 8.2), can request for a high-resolution screen nearby the shelf and display contents with product comparisons on it. The PM maintains a stack of presentations, and implements conflict resolution strategies to decide which content to present on which display (see Section 5.3.2). Another service of the instrumented environment relevant for this work is spatial audio service SAFIR, which allows to virtually position and render sounds in a room, when the smart object does not have integrated sound capabilities (more details in Section 5.3.3).

AmI Applications contain the domain specific part of an end-user application and are also able to trigger state transitions modelled by the SceneMaker. Domain specific application states based on input generated by smart objects are also handled in this layer, as for example storage conditions of medical products (see 8.1).

A tuple space provides a repository of tuples that can be accessed concurrently in parallel and distributed computing environments. As an illustrative example, imagine of a group of processors that produce data units and a group of processors that use this data. Producers post their data as tuples in the shared space, and the consumers can retrieve data from the space that they are interested in, which is also known as the blackboard metaphor. We are further able to divide the data space by associating a type to each message, such that we can achieve a stratified blackboard structure. The communication and coordination within this architecture is based on such commonly accessible tuple space, which provides an indirect interaction mechanism featuring the portability, heterogeneity, economy and flexibility needed for an intelligent environment. Processes post their information to the space as tuples (collections of ordered type-value fields) or read them from the space in either a destructive or non-destructive manner. In comparison to client-server designs, the anonymity and persistence of tuples offers a high degree of failure tolerance in dynamically changing environments [Johanson and Fox, 2002]. If a requested service is temporarily unavailable due to a network failure or reboot, the stability of the client process will not suffer. We have chosen the **EventHeap** server and API, which have been developed at Stanford University as a framework for their iRoom project [Borchers et al., 2002]. Similar implementations are available by Sun [Freeman et al., 1999] and IBM [Lehman et al., 2001].

The communication within this architecture can take place by point-to-point communication, but data exchange is more convenient via the EventHeap. The SceneMaker has been extended such that it can scan the tuple space for available sensors, smart objects or also services of the instrumented environment. In the same manner, AmI applications are able to receive raw data from sensors or already processed sensor data, such as interpreted user interactions. If interaction scenes involve the rendering of output through a display, the application sends output requests to the EventHeap, defining the contents and a target display, which may be part of a smart object or a shared device residing in the environment. After such output contents are delivered, the displaying system puts a confirmation message to the EventHeap, if required by the application. With this organisation of application and interaction flow control, the actual interface and application logic is only loosely coupled with hardware setups and allows to exchange e.g. the sensor instrumentation from a mixed arrangement (partially inside the object, partially in the environment) to a fully integrated one (all sensors inside the object) without any changes of the actual application software. Examples of AmI applications realised with this architecture will be described in Section 8.

5.3 Ambient Services

As previously outlined, smart object systems do not only involve functionality provided by object-embedded technology but also services of instrumented environments that are shared by several applications simultaneously. In the course of this work we have implemented and integrated several supporting services to assist in the implementation of smart object user interfaces as proposed in this dissertation. We will present three ambient services in the following sections: MaMUT (Macro and Micro User Tracking), a computer vision based approach of tracking multiple users in a room in order to determine which person is interacting with a given smart object system (also described in [Schmitz and Zimmer, 2007]); the Presentation Manager (PM) service that controls and allocates displays of the shared environment (as in [Stahl et al., 2005] and [Schmitz, 2006]); and the spatial audio system SAFIR (Spatial Audio Framework for Instrumented Rooms), for rendering sound in a three dimensional space in order to enable smart object systems without integrated sound output capabilities to generate appropriate auditory output (see also [Schmitz and Butz, 2006] and [Schmitz, 2003]).

5.3.1 MaMUT - Interaction-Tracking for Instrumented Environments

A common vision of UbiComp environments predicts that users are enabled to use a variety of services not only at home but also in public spaces, by interacting with the environment itself. One requirement of multi-user scenarios is the identification of users in order to provide confidential information or personalised services that take advantage of user profiles, histories and other data associated with a person. We attempted to design a system that requires users to identify themselves only once upon entering such an environment and subsequently allows to roam around and use devices and installations, while the system continuously recognises and tracks the user. Our computer vision-based approach disburdens the user from being equipped with further instrumentation like tags or personal devices, such as PDAs. A sort of identification is certainly required but only once at entering the environment and could be done in any way - e.g. with typing in a username/password combination or with a RFID tag. Another approach for user identification is realised by Feld et al., who introduce a system that exploits voice signals to both identify (if explicitly authorised by that person) and position a user within a car [Feld et al., 2010].

The instrumentation of interactive objects determines how it is used, but particularly shared public spaces mostly still lack the ability to detect, which of the sensed and present users triggered an interaction, if identification is not explicitly and repetitively done by the user. The service described in this section is a first step to solve this type of problem using off-the-shelf hardware and possibly even existing infrastructure, such as surveillance cameras.

The most similar application setting can be found in the EasyLiving [Krumm et al., 2000] project, Microsoft's intelligent environment, for which a very robust and sophisticated person tracking system was developed. The main difference is that they used expensive, non-standard equipment like stereo cameras. As one of our main restrictions was "runnable on off-the-shelf hardware", we could not build on this otherwise promising approach. Trackers that do run on off-the-shelf hardware are for example [Siebel, 2003] [McKenna et al., 2000] and [Cucchiara et al., 2005]. A successful operation of these systems in small sized intelligent environments like our instrumented room would not be feasible, but on larger scales such as underground stations or public parks. Interesting aspects of the last-mentioned paper were that the authors had to fuse data from multiple cameras and also passive infrared sensors in addition to *consistent person labelling*, which is also crucial for our system. A 3D person tracking system by Brubaker et al. is based on a model of lower-body dynamics, which characterises physical properties of bipedal locomotion such as balance and ground contact [Brubaker et al., 2007]. Ess et al. have realised a mobile vision system for person tracking, which is also able to track pedestrians in spite of frequent occlusions and motion of the camera [Ess et al., 2008].

On the other hand none of these trackers incorporates interaction detection, which is essential for our aims. Interaction detection is indeed integrated in a last tracker to mention, which is called W4S [Haritaoglu et al., 1998], and uses dynamic models of the human body to detect body parts. But the problem here was that, again, they could not work on off-the-shelf hardware. Overall, there exists a variety of approaches to detect and track movement of persons, for a more detailed and in-depth comparison of these systems the reader may refer to [Zimmer, 2006].

These approaches all have their strengths but cannot satisfy our needs, thus we decided to develop our own system to fill the gap between interaction detection and positioning utilising off-the-shelf hardware.

5.3.1.1 Tracking User Interaction

Associating user interactions detected by the environment with actual users can already be achieved by user tracking alone, if the position of all present users yield an unambiguous situation with only one user in direct vicinity of the device. But particularly in public spaces we cannot assume such a constraint, such that a second source of information (e.g. sensor technology) is required. We chose to use cameras as well, which has the advantage that users and devices do not have to be equipped with additional sensors and that we might further utilise existing infrastructure as for example surveillance cameras in malls or shopping centres. In our proposed system, cameras take over dedicated roles as either macro- or micro-tracker. For this purpose an appropriate positioning of these cameras is crucial: The macro-tracker should have an overview of a room and micro-tracker should observe an interactive device closely, which also means that each device that can be used by more than one user at the same time needs a dedicated micro-tracker. We also assume that a user is physically interacting with the interface, which holds for the majority of devices we have in mind, such as touch screen displays, installed keyboards and other TUIs like RFID-equipped products in a shelf.

Macro-Tracking: The actual macro-tracking part of our system tries to detect and position persons, which is basically done in three phases: First, we detect locations of motion in the images obtained from the macro-tracking cameras, then we try to identify persons in the detected moving regions and finally position each detected person by determining their room coordinates. The motion detection part uses a so called *background subtraction* algorithm, which keeps a model $m(x, y)$ of the background, i.e., the room without persons in it. For every camera image $f(x, y)$, we compute a binary motion image $d(x, y)$, classifying moving pixels. To reach this goal, we need a threshold parameter $T(x, y)$ (which may also be constant for all pixels), which tells us above which deviation from the model $m(x, y)$ we consider a pixel as moving. The actual algorithm we used is a sophisticated, adaptive statistical background subtraction algorithm using Gaussian mixture models, which is for instance described in [Stauffer and Grimson, 1999]. After we obtained our motion image, we first have to filter it, using a median filter, to remove small pixel regions which are most probably misclassifications of motion. The second part of the macro-tracking module consists of finding persons in the motion image $d(x, y)$. Therefore, we first do a *connected component labelling* on $d(x, y)$, which yields so called *blobs*, i.e. sets of connected pixels classified as moving. Unfortunately, several blobs will usually constitute one person, we therefore first assign a region (a rectangular bounding box) to each blob which has a minimal area, which means that we basically discard small blobs caused by noise. After that, we can merge regions in accordance to certain criteria, to obtain sets of blobs, forming a *person region*, also described by a rectangular bounding box. This process is inspired from [McKenna et al., 2000] and is illustrated in Figure 5.7.

The last step of macro-tracking is the positioning of the detected persons, where we used a *homography estimation* via the DLT-algorithm, as described in [Hartley and Zisserman, 2004]. We further assumed the use of static cameras and that persons will stand on the ground plane of our room, so we can position every person by the

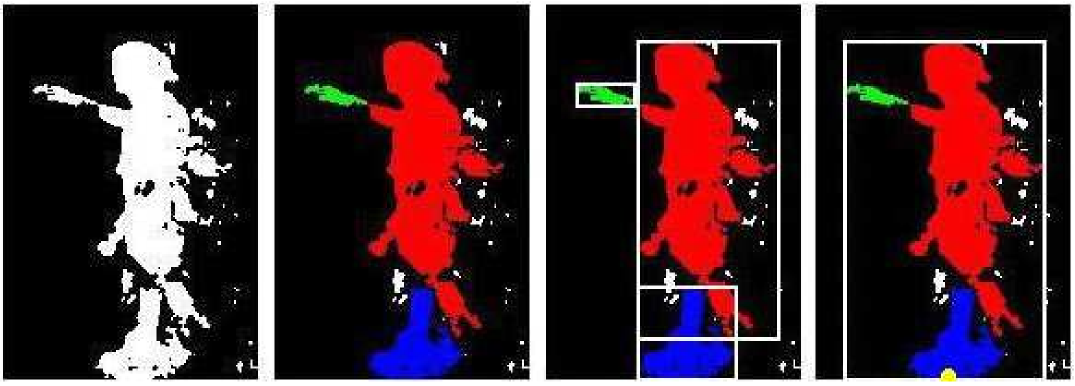


Figure 5.7: *Region merging. From Left to Right: Part of motion image. Connected component labelling yields three non-negligible blobs (coloured). Assign one region per blob. Merged regions to one person region with* tracepoint $\mathbf{x_t}$ *(yellow spot)*

(homogeneous) coordinates of its tracepoint $\mathbf{x_t}$ (cf. Figure 5.7). Now we have to map $\mathbf{x_t}$ from the image-coordinate system to $\mathbf{x'_t}$ in a world-coordinate system of our room. For this purpose the DLT-algorithm computes a homography matrix $\mathrm{H} \in \mathbb{R}^{3\times3}$ s.t. $\mathrm{H}\,\mathbf{x_t} \approx \mathbf{x'_t}$, which approximately transforms image- to world-coordinates. To compute H, the algorithm uses in our case just 4 point-to-point correspondences $\mathbf{x_i} \leftrightarrow \mathbf{x'_i}$ of points in the corners of the ground plane. One last question remaining is how we actually track detected persons (i.e., their regions) through consecutively captured frames from our macro-tracking cameras. The basic idea we use are *overlapping bounding boxes* (inspired by [McKenna et al., 2000]), which assumes that persons will not move too much during two consecutive frames from the cameras. Assuming that we have detected a person at time t, the position of its bounding box will be stored in a data structure in addition to other important person attributes like the tracepoint coordinates, a trajectory of former tracepoints, and an ID. If we detect a moving region with a bounding box that overlaps with the one of our person at time $t + 1$, we assume that this box corresponds to the same person and update the person's attributes according to the new measurements. Overlapping is simply checked by testing if the projections of the rectangular bounding boxes to the x-axis and the ones to the y-axis overlap. But what if a person did move farther away or if we lost their track for some frames, because of some disturbances such that their consecutive bounding boxes will no longer overlap? In this case our remedy (as well inspired by [McKenna et al., 2000]) builds on storing a *colour histogram* for each tracked person. If we now face non-overlapping bounding boxes, we test if the colour histogram of the moving pixels inside the box matches a stored histogram of a tracked person, and if so, we update the persons attributes by this measurements.

Micro-Tracking: This part of our problem space focuses on special areas of interest, e.g., a product shelf, and tries to detect user interactions with objects there, such that we mainly use *skin detection* to detect hand- and arm-regions. The actual skin detection approach we used is similar to the skin detection part of the *CAMShift Algorithm* [Bradski, 1998] and uses histogram back projection to finally obtain a binary skin image s_{bin}, which classifies skin pixels. To model a notion of skin colour, we use several training skin images, which are small images just containing skin, e.g. small patches cut out from micro-tracking camera images. These images are transferred from the RGB colour space to the HSV colour space. The major advantage is that the Hue channel carries all the colour information separated from the brightness information, which are contained in the other channels. Thus we use an accumulated and normalised H-histogram as a model of skin-colour, denoted $\mathcal{M}$, obtained from the sample skin images. Based on this, we compute a greyscale skin probability image s_{prob}, given the image from the micro-tracking camera f via $s_{prob}(x, y) = \mathcal{M}[f_{hsv}.h(x, y)]$, where $f_{hsv}.h$ denotes the H-channel of f transformed into the HSV colour space. A simple thresholding of s_{prob} yields the desired s_{bin}, where we finally form skin regions, corresponding to one detected hand or arm, which are the equivalent to person regions in macro-tracking.

Fusion of results: After macro- and micro-tracking are processed, we still have to fuse the gathered results by assigning detected hands (skin regions from micro-tracking) to detected persons (person regions from macro-tracking). The simple approach we follow, is to *cluster* skin regions, which are close together, whereby a cluster is just a set of skin regions. This idea is basically inspired from the clustering of blobs in [Krumm et al., 2000]. Our aim is now to obtain one cluster per detected person in the area of interest. From this one approach we can just ”order” the skin clusters and the detected persons, say from left to right (i.e., by increasing x-coordinates) and do a trivial one-to-one mapping from skin clusters to persons. A result of this process is illustrated in Figure 5.8.

Clearly, this can just be a proof-of-concept, it works for more than one person, but it has for example problems if persons cross their hands. We could most likely improve robustness by incorporating some probabilistic reasoning, as for instance described in [McKenna et al., 2000], where the authors were striving for a partial depth ordering of persons forming a group.

Certainly there is still work to do in different modules in order to improve robustness and reduce costs of installation and maintenance to create a system useful under real-world circumstances, e.g. including a more sophisticated background subtraction algorithm with shadow removal, as in [Cucchiara et al., 2005] or a more robust tracking approach, incorporating probabilistic ideas [Cucchiara et al., 2004]. An important issue is robustness, which is a problem in almost all Computer Vision approaches. For example, under unfavourable illumination conditions or if the room is too crowded with persons, our tracker will have severe difficulties. A second issue are the realtime needs, as we use some rather complex algorithms, which have to process several frames per second from several cameras, imposing hard requirements to the applied algorithms and on the used hardware.

5.3.2 The Presentation Manager - Flexible Management of Output Devices

As the price trend for large screens continues to drop, more and more public displays are being installed. Polymer and organic LED displays of arbitrary size are soon entering the market and are likely to be integrated into all kinds of objects at negligible cost. Today, public displays are typically poster sized and present product information, advertisements or transport time-tables at regular intervals. In the research lab, public displays haven tradition-

Figure 5.8*: Macro-tracking: Person regions with position in world-coordinates, moving direction and ID (top). Micro-tracking: Skin region associated to person with ID 2 (bottom)*

ally been used in the context of CSCW[2] projects. Whereas these projects use a single large display to share information between multiple workgroup members, we are more interested in how individual users can benefit from multiple distributed displays of all sizes.

Our instrumented environment employs public displays for a number of systems - including smart object applications - that provide users with shopping assistance and navigational aid. The key to such personalised and localised presentations is spatial knowledge about the environment, i.e. the position of the displays and the user. Given this information, the presentation service is able to present individual content on a suitable display nearby to the user.

Instead of running a single custom application on the public display, we favour all-purpose World Wide Web technology such as HTML and Flash for interactive presentations. This approach allows for a non-exclusive usage of the displays by multiple applications. Besides product information and navigational aid, the same display might also provide messages

[2]Computer Supported Cooperative Work

and postings that are maintained by completely different applications. In such a scenario of multiple users and applications, conflicting presentation requests are likely to arise and need to be resolved. Therefore the presentation service has to implement intelligent algorithmic strategies to allocate the most suitable display for a certain content and to divide the available displays by time and screen space, if necessary.

5.3.2.1 Managing Presentations in Instrumented Environments

In a public space with various displays we assume that a number of applications are running simultaneously and concurrently attempting to access display resources. A straightforward procedure for such applications would be to directly connect to the machines that control the displays - potentially creating conflicts with requests of other applications. In such a case each application would have to include its own routines to resolve conflicts. Our approach for the distribution of display resources is to insert a service between applications and public displays that hides the underlying display space with its possible conflicts and technical properties. The presentation management component resolves arising conflicts and distributes presentations in a context sensitive fashion (see screenshot in Figure 5.9). This centralised approach leads to a number of advantages and enables us to apply new concepts as described in the following section.

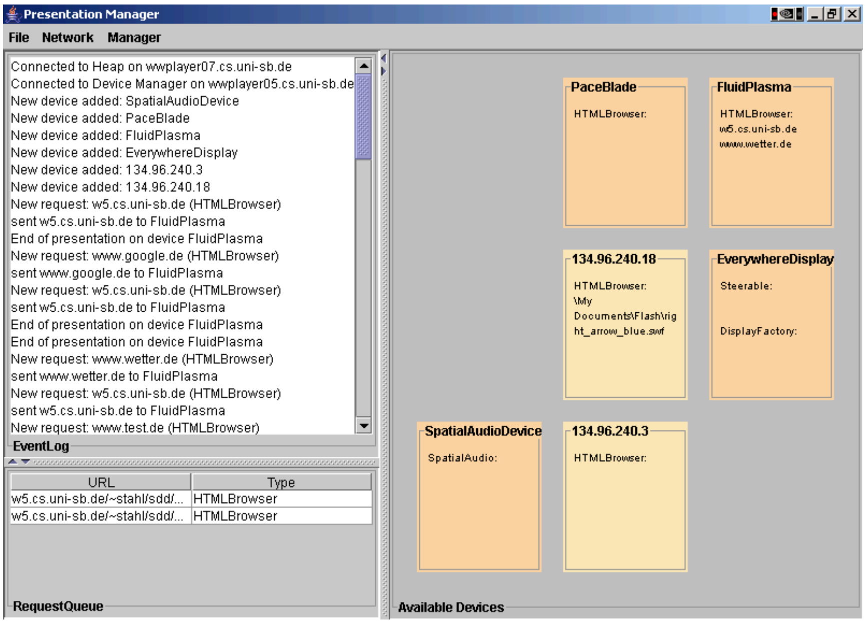


***Figure 5.9**: Screenshot of the PM monitor*

The PM is an abstraction of the public display space from an applications point of view and provides a high-level interface to issue presentation requests. Whereas canonical conflict resolution strategies could be first come, first served or priority-based assignments of resources decided to employ a set of rules for media location organisation: Presentation strategies are modelled as sets of filters and rules that are applied to a list of available displays and queued presentation requests. This algorithm generate schedules at runtime, which define how the presentations are distributed. We will elaborate a framework of this approach in the remainder of this section.

A presentation request sent to the PM must have the following syntax:

Source	URL of the media with the content
Destination	display or location or user
Type	image, text, video or mixed

Optional parameters can be provided to guarantee certain properties of the media rendering:

Minimum Display Time	e.g. 60 seconds
Minimum Display Size	small, medium, large
Minimum Resolution	e.g., 800x600
Audio Required	Yes / No
Input required	Yes / No

Based on this information the PM now organises the presentation in four steps:

1. *Generate a list of feasible displays:*
   This task is trivial if a particular display is given as the destination. If a person is the stated addressee of the presentation, first their location could be retrieved by e.g. the user-tracking service introduced in the previous chapter, whereupon the spatial location model is inquired to determine nearby displays, which could either be based on the euclidean distance using coordinate vectors or - more reasonable - a qualitative distance, i.e. user and display reside in the same room. Having exact locations of users and displays we are moreover able to examine the visibility of the presentation, considering the visual demands of its content in relation to the size, distance and viewing angle of the display, although we currently do not consider obstacles that potentially occlude the line of sight. All displays fulfilling these requirements now constitute the list of feasible displays for a given user.

2. *Sort the list*
   The list generated by step 1 is now sorted applying the criteria in following order:

   - idle displays
   - given minimum criterium (e.g. size)
   - general quality (resolution, size)
   - presentation duration

   The presentation will be displayed on the first available display on the current list.

3. *Resolve conflicts*
   If step 2 does not deliver any idle displays, we check if ending presentations will release

a display before the deadline of the current request elapses. Displays for which this condition holds are sorted according to step 2 and the request will be queued on the first display. If still the request cannot be scheduled (the latter list is empty), we check if a display can be shared - preserving the given requirements (resolution, size) of the involved presentations. If this does not resolve the conflict, we attempt to re-schedule currently playing presentations on different screens.

4. *Start over*
After applying these steps without success, we start over and consider more distant displays in combination with an audio notification, which is automatically added to the presentation by the PM.

This set of rules provides coherent presentations in public spaces and resolves conflicts by dividing display resources in time and space: Presentations are scheduled according to their deadline requirements and are delayed until resources are available. Screen space is shared if an appropriate device is available such that presentations are rendered simultaneously on the same screen in different windows.

5.3.2.2 Implementation

The PM is a component of the service layer, and permanently runs as a background task that listens for incoming requests on the *EventHeap*. Through this tuple space it also receives notifications whenever devices log on or off. Displays are also activated over a remote method invocation interface, which uses Active-X controls to access instances of the Microsoft Internet Explorer, managing their layout and contents. Display positions are matched with the user's range, and presentations are queued until displays become available or multiple presentations share a screen by opening multiple browser windows beside each other (division in time and space). The actual presentations on the displays are all done using a Web browser with Flash plugin, which is available for many platforms. A variety of authoring tools for Web-based media already exist and we are particularly able to reuse existing contents in this environment immediately. Moreover HTML already provides resource adaptive layout techniques that adjust to different screen and window sizes.

In our lab we have included different types of displays into the PM system, including a large wall-mounted screen, stationary panel PCs and PDAs, a table PC mounted to a shopping cart, and a steerable display. The SAFIR system described in Section 5.3.3 is handled separately, since acoustic messages can be delivered and perceived in parallel, such that this display can be shared at all times.

5.3.3 SAFIR - Spatial Audio for Instrumented Environments

Real-time 3D sound rendering has proven to be valuable across a wide range of domains and applications, including virtual reality, gaming and assistive interfaces for the visually impaired, as in e.g. [Mereu and Kazman, 1996]. We believe that use of spatial audio can also be beneficial in pervasive computing environments, particularly in smart object scenarios, in which objects do not contain any devices with sound generating technology. For these aims we have developed SAFIR, an API that allows for the easy, low-cost, flexible integration of spatial sound into UbiComp environments. SAFIR offers several benefits for use in pervasive computing environments over existing spatial sound libraries and APIs . Unlike many special-purpose spatial sound systems that use expensive hardware and proprietary interfaces (e.g. [Bargar et al., 1994]), SAFIR works with low-cost, off-the shelf hardware, and does not require developers to have DSP knowledge to install and configure spatial audio setups. Additionally, SAFIR is designed specifically to support the use of spatial audio in pervasive computing environments. Conventional spatial sound libraries that are built to support common surround standards and gaming applications under Windows operating systems assume that a user is stationary and oriented toward a single visual display. Such libraries also typically assume the use of a conventional 5.1 speaker setup, and therefore do not support flexible speaker configurations. SAFIR's sound output allows users to move and interact freely within an environment. As a cross-platform system SAFIR is not limited to Windows operating systems and furthermore allows for flexible configuration and re-configuration of speakers in the physical space to suit the purposes of the environment.

SAFIR was developed to meet the following design requirements:

- *Cost-efficiency:* The system should work with affordable, commonly available hardware.
- *User-centred sound:* The sound rendering should adapt to a moving listener position, correctly preserving the spatial image of the acoustic scenery as much as possible.
- *Configuration flexibility:* The system should support an arbitrary number and flexible spatial configuration of speakers in order to be able to deploy the same system in different setups for different purposes.
- *Ease-of-development:* UbiComp developers without digital sound processing experience should be able to use the system effectively.
- *Cross-platform compatibility:* The system has to be compatible with both Windows and Linux operating systems.

SAFIR works with almost any number of speakers. The minimum for any reasonable spatialisation is 4 speakers, but the maximum is only set by the limits of PC sound hardware. More speakers increase the quality of spatialisation. The volume spanned by the speakers defines the area in which sound can be freely spatialised. The developer defines the actual speaker positions in a configuration file when the system is set up. Generally, speakers should be distributed evenly in the listening area, but a higher spatial resolution in areas of particular interest can be achieved by positioning more speakers there. Developers use SAFIR by providing sound sources (audio files or streams) and their designated coordinates in the environment; other parameters such as volume or routes instead of simple coordinates are optional.

The VBAP (Vector Based Amplitude Panning) algorithm [Pulkki, 1997] is used for the general spatialisation, because it provides computationally cheap functions to simulate the direction of a virtual sound source. In a 2D setup (i.e., where speakers are placed in the same plane), VBAP selects two speakers and places the virtual sound source between them by adjusting the respective gain values for the two speakers. Figure 5.10 visualises how two loudspeakers are used to render a virtual sound source to realise a talking product in a shelf. The vector p points at the talking product and is decomposed into the vectors $g1l1$ and $g2l2$, which point at the neighbouring loudspeakers and define the actual gain parameter of the respective loudspeaker. As a result, a virtual sound source is created at the position of the talking object, which is perceived as the physical source of the sound.

In a 3D setup, three speakers are chosen to interpolate a direction within the triangle spanned by them. Since VBAP only simulates the direction of a sound, we added additional mechanisms to create a feeling of distance and moving sound sources as well. The main distance cues are the decreasing intensity of a sound with increasing distance due to air absorption, and the increasing delay due to the limited speed of sound. The system therefore attenuates virtual sound sources in proportion to the square of their distance and delays them proportionally to their distance. As a side effect, the changing delay of a moving sound source correctly creates the Doppler effect, 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. If the user position can be determined (e.g. via a person tracker as introduced in Section 5.3.1), the coordinates can be passed in to the spatial sound engine, which will adjust the synthesis of currently playing sounds to it, which especially improves the spatial perception significantly when the user moves around a virtual sound source.

We deployed SAFIR in the SUPIE at a hardware cost of approximately $1000 (for 8 speakers). It is used to support interactions in a variety of scenarios: For instance, users engage in dialogues with physical objects which often cannot produce audio on their own. The spatial audio display described in this section supports the anthropomorphic object metaphor by providing audio channels that can be spatially related to objects in our instrumented environment.

5.4 Synopsis

The prototyping testbed introduced in this chapter implements several key features that regard the particular requirements of realising TASOs in instrumented environments:

TASO Platform: A modular overall architecture for flexible integration of hardware and services embedded into objects and shared by the environment, integrating a toolkit for finite-state-machine based modelling of anthropomorphic states in smart objects.

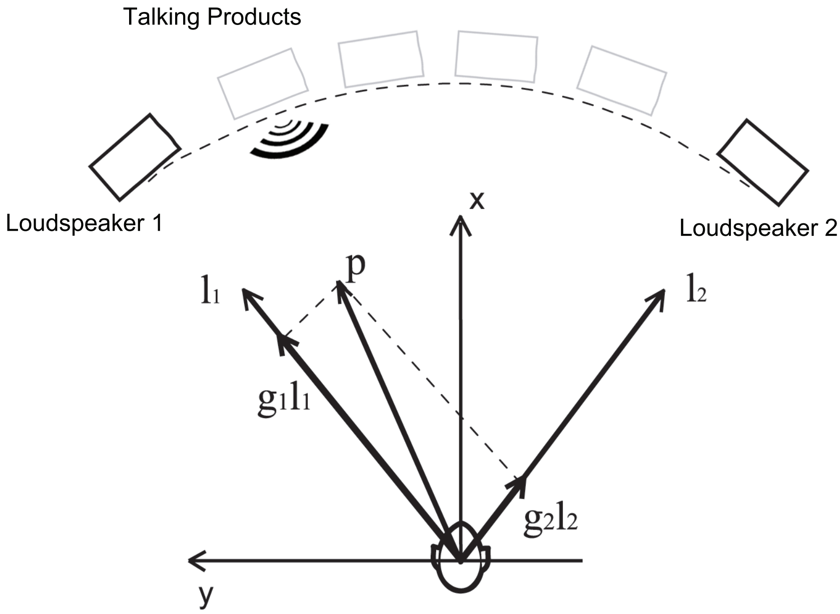


***Figure 5.10**: Realising auditory displays for talking products with SAFIR*

TASO SensPro: A library and graphical user interface for sensor-based prototyping of smart objects, which enable rapid development with reusable code fragments in stream processing structure.

MaMUT, PM and SAFIR: Ambient services that support smart object applications with additional functionality provided by the instrumented environment.

The loosely coupled, modular architecture reduces the amount of knowledge about potentially complex interfaces to services, distributed hardware and other components that applications would need to know to integrate them. Instead, a unified messaging exchange via a asynchronous tuple space facilitates all communication between prototyping components. Furthermore, interaction recognition can be fused from various sources, which might reside inside smart objects or also be situated in the instrumented environment. Several applications can share the same input and output resources in parallel, which includes sensors as well as displays. The integration of the SceneMaker provides developers with a flexible toolkit to model anthropomorphic interaction with smart objects in a rapid prototyping approach.

We further described a framework for sensor-based prototyping, which helps to quickly develop software components that rely on various sensor data streams. Different types of classes for sensor stream processing modules (*input, output and filters*) implemented and provided as a Java library, which we have named TASO SensPro allows researchers and developers to assemble applications without the burden of investing much time into low level sensor stream processing. The library can be easily extended for upcoming tasks, which facilitates a framework to afford future reuse of sensor-based code fragments by simple means.

Several ambient services hosted by our instrumented environment are introduced as flexible supporting building blocks in prototyping endeavours. MaMUT applies computer vision concepts for user tracking to enable the association of interaction events to users, enabling applications in intelligent environments to provide personalised services in multi-user scenarios, e.g. in public spaces. The PM for intelligent environments is a centralised service that provides other services and applications with an abstraction of the available display space within the intelligent environment. The PM allows the design of applications without bothering about the details of display allocation. We have elaborated on different conflict resolution strategies to find appropriate display devices in a specific situation. SAFIR was briefly introduced in this chapter and constitutes a general purpose auditory display to enable smart objects without sound capabilities to make use of the acoustic channel with for instance speech or anthropomorphic sounds and to spatially associate the sounds with physical objects.

6 Vocal Behaviour Generation for Anthropomorphic Objects

The creation of TASOs as described in this chapter requires novel approaches to express affect and personality in acoustic interfaces. For this reason we have tackled several key aspects of TASOs that have not been explored yet. We will start off with describing the concept of *anthropomorphic auditory icons* in Section 6.1, which helps to create convincing animalistic objects through the use of anthropomorphic sounds.

For the expression of personality of talking objects we modelled the prosodic parameters of their speech [Schmitz et al., 2007b] [Schmidt, 2005], which will be discussed in Section 6.2. Complementary, the idea of an *anthropomorphic base grammar* introduced in Section 6.3 discusses an approach and guidelines to phrase utterances in several variations that reflect different personalities, which a smart object can chose from during runtime (see also [Schmitz et al., 2008a] [Hollinger, 2007]).

6.1 Anthropomorphic Auditory Icons

Sound can be used in user interfaces for several kinds of information, for example as immediate feedback to users' actions, such as clicking a button or to inform the users of the state of internal computer processes. Sound elements in user interfaces exhibit a number of beneficial attributes, for instance that users can focus their attention on a particular sound and be aware of several other sounds simultaneously. Further, continuous sounds can fade into the background when not actively attended and the acoustic attentional focus can be shifted between several sound concurring sources.

One popular type of sounds in user interfaces are the so called *auditory icons*, which are everyday sounds designed to convey information about events by analogy to everyday sound-producing events and thus, can be intuitively interpreted by users (see [Jung, 2009] for an overview of non-speech auditory interfaces). For example, if a user thrashes a text file on a computer desktop, the sound of paper being crumpled may be rendered to support informing the user of the successful execution of the action. The type of the object that the users hears might indicate the type of file that has been deleted, while the file size might be indicated by the size of the object being acoustically destroyed. Such auditory icons are used like sound effects on desktop computers, supporting visual output with appropriate sounds, which allows users to listen to computers as to the everyday world. Creating auditory icons follows a similar strategy as in the creation of visual icons, since their appearance draws on real world analogies and their interpretation relies on prior knowledge and experiences of the user.

In this section we will develop a novel type of auditory icons, which we will call *anthropomorphic auditory icons*. This type of auditory icons exploits the distinct ability of human beings of interpreting sounds that originate from living beings, such as human affect bursts or physiological sounds. Such anthropomorphic auditory icons can be employed by animalistic objects to reflect basic needs and desires that relate to, for example, pain, hunger or desire for sleep, which correspond to application specific computational states. This type of acoustic feedback and reflection of internal states enables simplistic smart object interaction as outlined in Section 4.3.1.

Research regarding emotions and sounds in general has been carried out in various disciplines, mainly studying the emotional impact of different kinds of acoustic stimuli on the listener. Bradley et al. for instance have observed the affective responses to everyday sounds [Bradley and Lang, 2000]. For the discipline of auditory icons it has been of particular interest how a certain level of alertness can be achieved, for instance in driver-vehicle interaction scenarios [Larsson, 2010] or medical contexts [Edworthy and Hellier, 2006]. In the following we will briefly examine previous work in the field of auditory interfaces in smart object interfaces and approaches to anthropomorphic sound generation, before we determine existing anthropomorphic sounds from the real world and finally describe our approach of creating anthropomorphic auditory icons for smart objects.

6.1.1 Smart Objects and Sounds

Research about auditory displays in the domain of smart object systems is scarce, there are only few structured attempts that particularly regard this discipline.

The 2-years EU project *"The Sounding Object"* (SOB) was funded within the Disappearing Computer Initiative and examined sound models that are based on the physics of sound-generating phenomena for the integration within artefacts and appliances that interact physically with humans (see e.g. [Rocchesso et al., 2003]). The project aimed at developing sound models that are responsive to physical interactions and that can be easily matched to physical objects and constitutes an alternative approach to traditional signal-based sound design (as e.g. in sonification). These physics-based sound models that can be parametrised are supposed to provide basic blocks for sound generation, as they exhibit natural and realistically varying dynamics. One implementation example within this project was the virtualisation of an irish percussion instrument, the so called *bodhran*, which is a frame drum played with a double-sided wood drumstick. The research team of the SOB project instrumented these two artefacts to reproduce the timbral characteristics and to provide access to all its controlling parameters such as resonance frequency, damping, mass of the drumstick, and impact velocity.

While such physics-based sounds could theoretically complement the tangible interaction part of our concept, it does not regard the anthropomorphic notion that we strive for. More closely related research can be found in the realm of human-robot interaction, as for instance conducted by the group around Takanori Komatsu at Shinshu University, Japan. Komatsu has tried to express what he calls *attitudes* of artefacts through beep sounds, which is basically a simplified variation of anthropomorphic feedback. He distinguished *positive*, *neutral*, and *negative* attitudes, which can be used to express *agreement*, *hesitation* and *disagreement* in human-object interaction [Komatsu, 2006]. In order to examine whether such attitudes can be assigned to artefacts in subtle manner through simple beep-like sounds, he carried out a user study in participants where requested to listen to beep tones that vary in duration and pitch progression. The results revealed that sounds with rising tones are generally perceived as negative/disagreement regardless of their duration and that flat sounds with relatively longer duration were interpreted as neutral/hesitation. Furthermore, falling tones with shorter duration were perceived as positive/agreement. Interestingly, Komatsu et al. found in follow-up experiments that the agents' appearances affected people's interpretations of the agents' attitudes, even though the agents expressed the same information [Komatsu and Yamada, 2008]. Most misinterpretations were false negative ratings of positive and neutral sounds. The researchers hypothesised that the relationship between the agent's appearance and the expressed sounds were unfamiliar to the participants, such that these sounds were interpreted as indicating that something wrong had happened, which is a common interpretation of beeping sounds. This would imply that agents should *"express imaginable or predictable expressions in order to inform people of their attitudes"* [Komatsu and Yamada, 2008].

Various examples of anthropomorphic sounds that come close to what we envision to apply in animalistic smart object interaction can be found in professional movie productions. Mickey Mouse and Donald Duck were early examples of anthropomorphised animals, which communicated both verbally and through affective sounds. Further popular examples are the audifications of the robot R2D2 in the Star Wars movie series and of Wall-E, a cartoon robot in a movie with the same name. R2D2 is basically a talking robot that does not speak human language but instead combines beeps and whistles to express itself. Thus, these sounds had to convey a basic sense of what the robot is trying to express, mostly affective reactions such as being angry, sad or rude towards his robot friend C3PO. Wall-E "talks" quite similarly, except that he is able to utter names of a few principle characters he encounters, and that he is able to vocalise in terms of making grunts and moans, therefore electronically mimicking human affect bursts (see Section 6.1.2).

The sound designer of both movies, Ben Burtt, has remarked that some of these sounds were purely made on a synthesiser, while others origin from recorded mechanical and motor sounds. Since all these sound effects were created ad-hoc by an expert for their intended use, there is yet no structured and reproducible methodology to reconstruct similar anthropomorphic sounds for other contexts and use cases.

6.1.2 Affect Bursts and Physiological Sounds

There are several examples of work that examined non-speech sounds emanated by living entities. One major category are the so called *affect bursts*, which was introduced for the first time by Scherer [Scherer, 1994], defining them as *"very brief, discrete, nonverbal expressions of affect in both face and voice as triggered by clearly identifiable events"*. Schröder has redefined this term for his work as *"short, emotional non-speech expressions, comprising both clear non-speech sounds (e.g. laughter) and interjections with a phonemic structure (e.g. "Wow!"), but excluding "verbal" interjections that can occur as a different part of speech (like "Heaven!", "No!", etc.)."* [Schröder, 2003] and evaluated affect bursts produced by actors expressing different emotions. In his studies he collected a set of affect bursts and grouped them into 10 emotion categories, namely *admiration, threat, disgust, elation, boredom, relief, startle, worry, contempt and hot anger.* A mean recognition score of 81% by human subjects indicates that affect bursts seem to be effective means of expressing emotions even without context.

A similar approach was taken by Pelin et al. who have recorded and validated a set of affect bursts, which they called the *Montreal Affective Voices* and made available for download [Belin et al., 2008]. This set consists of 90 nonverbal affect bursts corresponding to the emotions of anger, disgust, fear, pain, sadness, surprise, happiness, and pleasure, recorded by 10 different actors. The researchers also collected ratings of valence, arousal, and intensity for these eight emotions from 30 participants. Their analyses revealed high recognition accuracies for most of the emotional categories (with a mean of 68%) and also significant effects of both the actors' and participants' gender: The highest hit rates (75%) were obtained for female participants rating female vocalisations, and the lowest hit rates (60%) for male participants rating male vocalisations. The selection of emotions was done in accordance to the widely used *Ekman Faces* [Ekman, 1993], which states that the emotions anger, sadness, fear, surprise, disgust, contempt and happiness have distinct facial expressions. These emotions were extended with pain and sensual pleasure. To complete the compilation of life-like sounds for anthropomorphic auditory icons, physiological sounds have to be considered as well, which represent organic states or activities, such as sounds of swallowing or groaning.

6.1.3 Generating Anthropomorphic Auditory Icons

We aimed at finding the means to extract the attributes of aforementioned affect bursts, which are responsible for the recognition of affect and to apply them to new *carrier sounds*, which could then be associated to virtual or physical entities such as smart objects. This would enable us to flexibly generate new affect bursts for arbitrary sources by designing carrier sounds that reflect the peculiarities of that object and by modulating it to express a corresponding affective state. The optimal solution in theory would be to remove all human-like acoustic attributes from recorded affect bursts, but at the same time keeping sufficient acoustical information that would enable a listener to perceive the intended affective notion with the same efficiency as with the original sound material.

In order to systematically synthesise anthropomorphic auditory icons, we started with an analysis of the recordings of the Montreal Affect Bursts. Informal listening tests and an analysis of the spectrograms of the given sound samples indicated that it will be hard to extract the emotional component from several affect bursts, which we suspected to be grounded in subtle nuances beyond pitch and envelope. We therefore recorded an additional series of affect bursts in a recording studio with a 27 years old male speaker and aimed at creating a wide variety of affective expressions with which we can conduct further experiments to identify the most suitable material. In total we have recorded 97 affect bursts and 43 samples with different kinds of physiological sounds. Figure 6.1 provides an overview of the categories of the recorded sound samples; the numbers in brackets display the amount of sound samples in the respective category. These recordings (plus the Montreal Affect Bursts) could directly be incorporated into realisations of TASO systems or, as described in this section, provide a basis for further synthesising efforts.

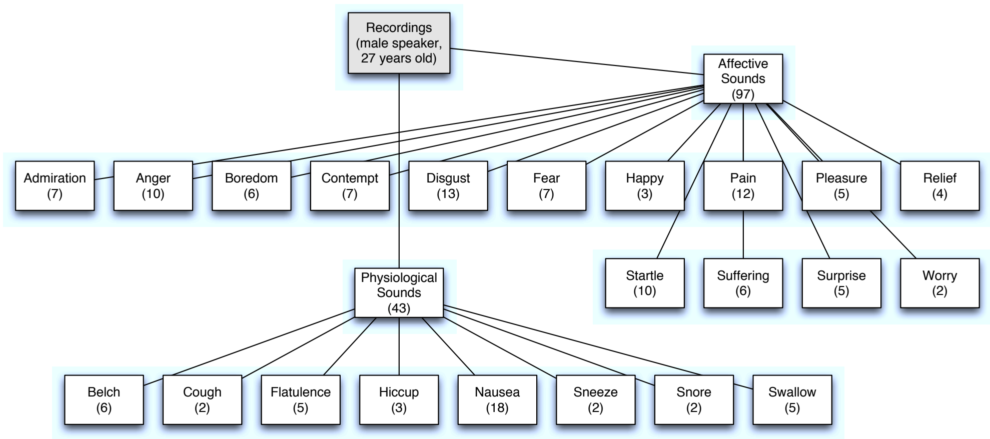


***Figure 6.1**: Overview of newly recorded and classified sound samples*

A literature review in the field of electronic sound engineering lead us to a so called *vocoder*[2], which is recently more common in the domain of electronic music production. A vocoder applies a multiband filter on the *carrier sound* and the *modulating sound*, thereby dividing both source samples into a variable number of frequency bands. Furthermore, the vocoder follows the envelope of the modulating sound in each band and applies this envelope to the corresponding band of the carrier sound, which is realised through a Voltage Controlled Amplifier (VCA). Figure 6.2 displays a schematic diagram of this process. We utilised this method in order to extract the energy and pitch characteristics of the modulating sound (the original affect burst in our case) and to apply it to a carrier sound (for which we used dif-

[2] `http://en.wikipedia.org/wiki/Vocoder`, last visited June 30, 2010

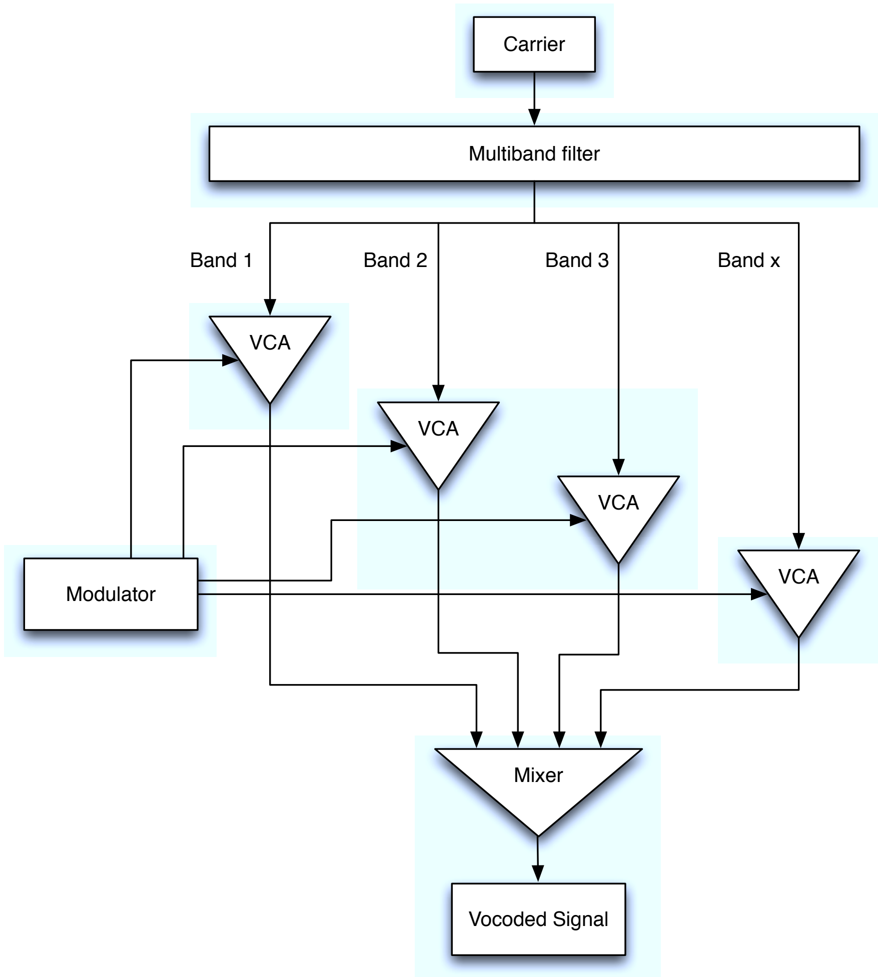


***Figure 6.2**: The vocoding principle*[1]

ferent artificially generated sounds, e.g. sine waves), which determines the overall timbre of the resulting sound sample. In addition to vocoding we applied custom optimisations on certain sounds, depending on the acoustic attributes of the particular affect burst. For example, for vowel-deprived sounds (like a snarl-like sound expressing *anger*) we added a low frequency oscillator to accentuate the rhythm of the sound in certain frequencies. Another method of amplification was to exaggerate the pitch curve afterwards by manually adjusting the pitch over time or adding a frequency shifter to create frequency vibration (as e.g. with the sound expressing *disgust*). Figure 6.3 displays waveforms and spectrograms of the original Montreal Affect Burst recordings of happiness and anger performed by speaker number 59, followed by the vocoded versions, based on a sine wave and a sawtooth carrier sound.

[2]adapted from `http://www.dma.ufg.ac.at/app/link/Grundlagen:Audio/module/8086`, last visited October 20, 2010

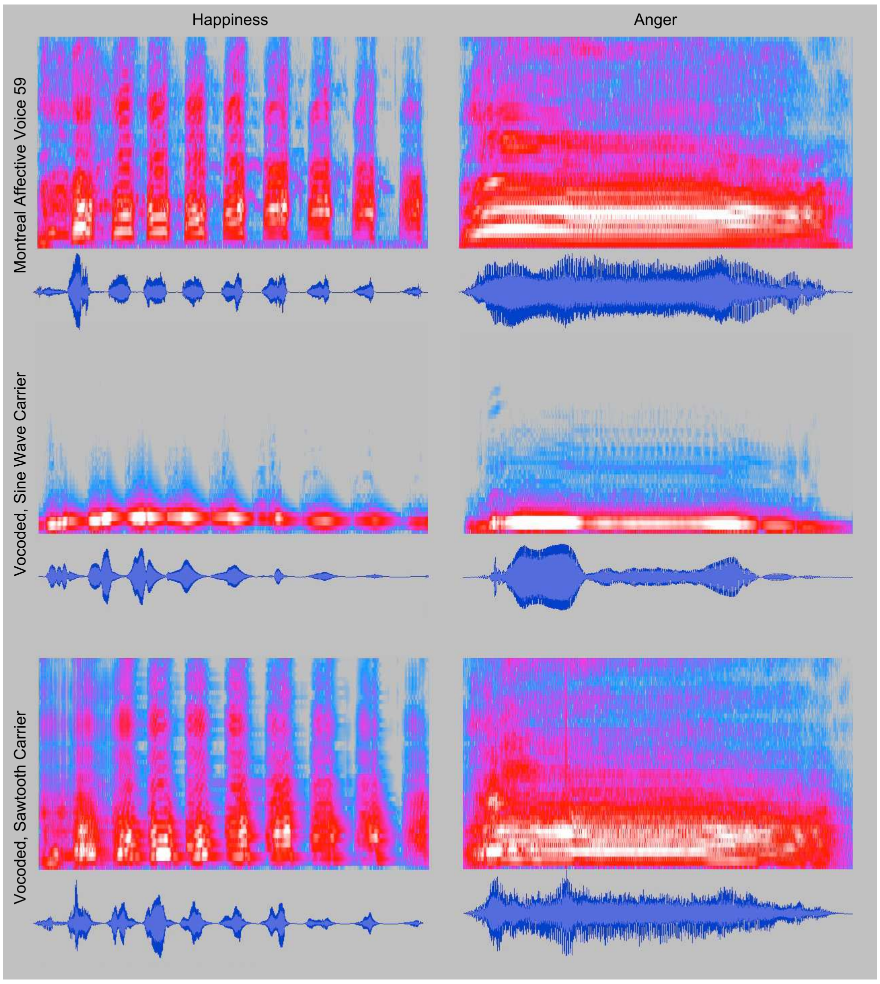


Figure 6.3*: Exemplary waveforms and spectrograms (0 - 8000 Hz).*

The sound representing *happiness* basically consists of laughter, resulting in a distinct energy pattern, which can be recognised in both types of diagrams and vocoded signals. Thus, the rhythm of the laughter is also recognisable with a sine wave carrier, which actually covers only a limited bandwidth. The opposite case can be seen in the *anger* example, which is a growl-like sound. The waveform does not have a very sharp profile, but a rather complex energy distribution over different frequencies. The diagrams show that the sine wave carrier is not suitable to reflect this, whereas the application of a sawtooth-based carrier achieves a more similar pattern of energy distribution over the whole frequency range. Since this is a consequence of the vocoding principle, a general rule for the construction of carriers must be to create sounds that cover a wide frequency spectrum.

As with these two examples, we have used *Ableton Live 8.1* and the *Eiosis ELS 22-band Vocoder* plugin for creating a a set of compound filters based on vocoding for each affect category that is performed by the Montreal Affective Voices, such that we have achieved a general framework to create anthropomorphic auditory icons for the categories *anger, disgust, happiness, pleasure, sadness, surprise, fear, and pain*. These filters can be applied to new carrier sounds to create anthropomorphic sounds customised to particular objects. In Section 7.1 we present a user study on the perception of affect in such synthesised affect bursts.

6.2 Expressing Personality in Voices of Talking Objects

The paradigm of talking objects with personality as introduced in Section 4.3.2 involves natural language dialogues with smart objects. We have reviewed related studies, which state that conversational interface agents with consistent personalities yield a series of advantages, such as an increased perceived usefulness, more fun, a lower cognitive load and enabling the user to memorise spoken contents easier. While there are several methods of implementing personality in the course of a dialogue, using the voice itself as a carrier of personality cues has the advantage that an impression of personality can already be conveyed with the first utterance, which is particularly useful in scenarios, where anthropomorphic objects are encountered only for very short interaction sessions. This also applies, for instance, to shopping use cases, which our group is working on for several years now and for which we have also realised a prototype in the context of this dissertation (see Section 8.2). In a shopping domain the product manufacturers would benefit as well from objects that express a sense of personality, since the personalisation of the product provides a new channel to communicate a brand image or distinct attributes of a certain product. A study within the context of marketing research showed that if in radio advertisements a voice fits the product, it helps the listener to remember the brand, the product and claims for that product [North et al., 2004].

In this section we present how we create personalities of talking products by modelling the prosodic parameters of their speech. We will first briefly introduce common personality concepts and review how personalities are reflected in speech.

6.2.1 The Big Five Personality Dimensions

In psychology the dimensions of personality are described by the five factor model [McRae and John, 1992] or variations. In parallel there are also models describing the personality of products and brands [Aaker, 1997] by identifying five dimensions as well, with three being similar to the dimensions of human personalities, as depicted in Table 6.2.1.

Human Personality Dimension [McRae and John, 1992]	**Brand Personality Dimension [Aaker, 1997]**	**Typical Attributes of Brand Personality**
Extraversion	Excitement	spirited, imaginative, daring
Agreeableness	Sincerity	down-to-earth, honest
Conscientiousness	Competence	reliable, intelligent
Neuroticism	Ruggedness	outdoorsy, tough
Openness	Sophistication	upper class, charming

Although these personality models are not equivalent, we can find the similarities between the dimensions of human and brand personality when looking at these terms within the hierarchical taxonomy based on the statistical technique of factor analysis (see [Schmidt, 2005] for more details), regarding the dimensions extraversion - excitement, agreeableness - sincerity, and conscientiousness - competence.

For our investigation of talking products in the shopping domain the personality model for brands is of particular interest for us. In a shopping context, personalities could obviously not only be assigned in order to discriminate *brands*, but also different *types* of products (e.g. plants, tools, sports equipment), or different *subcategories* of the same product type (e.g. business and outdoor mobile phones).

In the following we will describe a novel approach to support the personality of a product through the adequate choice of voice.

6.2.2 Personality and Voices

Very few researchers investigated how the exact attributes of brand personalities are reflected in voices. We therefore considered related more general research on voice modelling as well. We present some of our findings in this chapter to exemplify our literature review as the basis of the voice parametrisation: Excitement or arousal can be transmitted by a high mean pitch and high pitch variation [Laukka et al., 2005]. Furthermore amplitude and speaking rate are increased when excitement of the speaker is high [Krauss et al., 2002]. Speaking rate also has a strong effect on how competent a voice is perceived: Faster speakers appear more convincing, more intelligent and more objective [Smith et al., 1975, Apple et al., 1979]. Less pauses and repetitions and a more dynamic voice are assessed as more competent [Zuckerman and Miyake, 1993]. It has also been shown that louder voices and male voices are considered to be more logical than softer voices and female voices. We associated sincerity to benevolence, which is expressed by speech with more variable intonation [Brown et al., 1975], whereas high pitched voices sound less benevolent [Apple et al., 1979] and those with a high fundamental frequency more deceiving [Streeter et al., 1977]. In order to express sophistication we focused on its facets softness and attraction, which are associated with voices that are less monotonous, with a low pitch and also a larger pitch range [Zuckerman and Miyake, 1993]. Ruggedness, interpreted as robustness, can be achieved by a low, slow and loud voice, which is underlined by the so called frequency code [Ohala, 1994], describing a universal relationship between the size of a speaker/animal and the fundamental frequency of its voice.

We generated different versions of a neutral test utterance with the German speech synthesiser Mary [Schröder and Trouvain, 2003] and the MBROLA voices de6 (male) and de7 (female) [Dutoit, 1997]. We changed the four prosodic parameters pitch range (in semitones), pitch level (in Hz), tempo (as a durational factor in ms) and intensity (soft, modal or loud) according to our findings outlined above. Table 6.2.2 provides an overview of the prosodic parametrisation:

Personality Dimension	Pitch level	Pitch range	Tempo	Loudness
Baseline	0	4 st	+15%	Modal
Sincere	0	8 st	+15%	Loud
Excited	+30%	8 st	+30%	Loud
competent	-30%	8 st	+30%	Loud
sophisticated	-30%	8 st	+15%	Modal
rugged	-30%	4 st	+0%	Loud

Table 6.1: *Parameter matrix modelling the brand personality Dimensions.*

The parameters were either set to a higher or lower level in relation to the baseline or left unchanged, such that we have clear and distinguishable effects in each parameter change. Since the baseline voice was rather slow, we increased the default speed by 15%. This personality-parameter mapping can be used by the interface developer if the voice is supposed to express a high degree of a distinct personality dimension, in order to, for instance, match a chosen brand image. For an explorative user study (see Section 7.2) we generated in total 12 audio files: For each gender five distinct personalities plus the baseline.

6.3 Anthropomorphic Base Grammar

As seen in the previous section, we have created voices that reflect certain personalities according to Aaker's brand personality model [Aaker, 1997] only by adjusting prosodic parameters. We chose this model over the five factor model [McRae and John, 1992] commonly used in psychology, since we are applying the concept of talking objects in the shopping domain (see Section 8.2) in a first prototype and generally consider the retail arena as a major use case environment. However, both models are rather similar and to a certain extent exchangeable. The study in Section 7.2 shows that there are clear preferences for our prosody-modelled speech synthesis for certain brand personality dimensions. But not all personality dimensions were perfectly perceived as intended, such that we have to amplify the effect.

The personality model by Costa and McRae [McRae and John, 1992] (as introduced in Section 6.2.1) constitutes five dimensions of human personality: *Extraversion, Agreeableness, Conscientiousness, Neuroticism* and *Openness* on a scale from 0 to 100. Obviously, differentiating 100 discrete levels in one dimension is beyond the scope for our aims, therefore we simplified this model by discriminating three levels in each dimension:

- low: value between 1 and 44 (31% of population)
- average: values between 45 and 55 (38% of population)
- high: values between 56 and 100 (31% of population)

Related work, e.g. by Andre et al. [André et al., 2000] limited their personality modelling to only two of the five dimension - namely extraversion and agreeableness - since these are the most important factors in interpersonal communication. Nevertheless, in our literature study we discovered considerable influences of openness and conscientiousness to speech, therefore we incorporated these two dimensions as well. The effect of the dimension neuroticism is mainly to describe the level of susceptibility to strong emotions, both positive and negative ones [Costa and McCrae, 1985]. It was further shown that the level of *neuroticism* is very hard to determine in an observed person [Gill et al., 2006], thus we decided that four dimensions will suffice for our work.

We conducted an exhaustive literature review on how speech reveals different personality characteristics. Among numerous other resources, two research papers provided essential contributions to our work: Pennebaker and Kings analysis in *Journals of Personality and Social Psychology* [Pennebaker and King, 1999] and Nowson's *The Language of Weblogs: A Study of Genre and Individual Differences* [Nowson, 2006]. In both studies a large number of text blocks were examined with an application called *Linguistic Inquiry and Word Count*[3] (LIWC), which analyses text passages word by word, comparing them with an internal dictionary. This dictionary is divided in 70 hierarchical dimensions, including grammatical categories (e.g. noun, verb) or affective and emotional processes. Pennebaker determined in a study the 15 most reliable dimensions and searched for them in diary entries of test persons with LIWC (see Table 6.3). With these results together with the given personality profiles of the participants (according to the five factor model), he identified correlations between the two. Nowson performed a similar study and searched through weblogs for the same LIWC factors.

LIWC Factor	**Examples**
Articles	The, a
Causation	Because, reason
Discrepancies	Should, could
Exclusive	Without, but, except
First-person singular	I, me, my
Inclusive	With, and
Insight	Understand, realise
Negative emotion	Hate, envy
Negations	No, not, can't, never
Past Tense	Went, made
Positive Emotion	Happy, love
Present Tense	Go, make
Social	Friend, buddy
Tentativity	Perhaps, maybe
Words of more than 6 letters	Agreeableness, conscientiousness

Table 6.2*: The most reliable LIWC dimensions [Pennebaker and King, 1999]*

[3] `http://www.liwc.net/`, last visited March 27, 2010

Based on these results, we provided a set of recommendations how responses of an talking entity with a given personality should be modelled, particularly considering smart object scenarios. These recommendations can be divided into three categories:

Dialogue Behaviour (DB): This category regards the general behaviour of the object, e.g. is it pro-active in conversations or rather passive? Short-spoken or longwinded?

Speaking Style (SS): Recommendations of this category influence the overall construction of phrases, such as whether utterances should be more colloquial or formal.

Lexical Choice (LC): In natural language generation the lexical choice decides which content words (nouns, verbs, adjectives, adverbs) for a certain concept are selected in a generated text.

For instance, for a high level of extraversion these recommendations are given:

- Immediately greets the user upon contact, such as touching it (DB)
- Comparably more elaborate replies (DB)
- Frequent use of terms from a social context or describing positive emotions (SS)
- Usage of colloquial phrases, based on verbs, adverbs and pronouns (SS)
- Avoidance of *maybe, perhaps* and extensive usage of numbers (LC)
- Preferred bigrams: *a bit, a couple, other than, able to, want to, looking forward* and similar ones (LC)

The complete set of phrasing recommendations is summarised in Appendix A.2. Following these principles we implemented basic product responses (greetings, inquiries for product attributes, farewell) for several personalities. All possible replies are stored in one (XML-)file, which we named the *Anthropomorphic Fundamental Base Grammar*. All entries include an associated personality profile, for example:

```
<reply
    query="hello"
    reply="Hello, nice to meet you!"
    ag="1" co="2" ex="1" op="1">
<\reply>
```

Which means that this is the greeting of a product with average agreeableness, extraversion and openness and a high value in conscientiousness. Another example:

```
<reply
    query="hello"
    reply="Hi! I'm sure I can help you! Just tell me what
        you need and I bet we can figure something out!"
    ag="2" co="2" ex="2" op="2">
<\reply>
```

All entries that do not regard any particular personality, should have average personality values in all dimensions. A central product database with all products and their attributes is extended by the assigned personality profile, i.e. the values in each of the four dimensions. When the application starts up, it retrieves the product data of each product instance and extracts the appropriate entries from the base grammar to build the custom product grammar. If there are no entries that exactly match the given profile, the one that has the most identical values will be chosen. It is also easily possible to integrate the base grammar into a SceneMaker-based application by including the different utterances into the pre-scripted scenario descriptions and creating dependencies with variables that represent the object's personality (see also Section 5.2). A specific application generates a consistent speech interface to a product by knowing its attributes and a given personality profile, for instance preset by the manufacturer.

6.4 Synopsis

We have developed three novel methodologies to support the expression of personality and emotion, which can be leveraged as building blocks for the creation of different types of TASO interfaces. This includes the construction of so called anthropomorphic auditory icons, a novel approach of creating auditory icons that represent affective states acoustically in order to enable the expression of system states or feedback in user interfaces. This is the first conceptualisation of previous attempts of isolated solutions trying to express affective states in synthesised, non-verbal acoustic feedback. An initial evaluation of this approach will be presented in Section 7.1. We also recorded an exhaustive set of affect bursts plus physiological sounds, which can serve as complementary input for the creation of anthropomorphic and animalistic interfaces.

We have further introduced a new approach of expressing personalities of talking products by modelling the prosodic parameters of their speech, which is examined in a user study in Section 7.2 as well. A more detailed discussion about how these parameters are derived from related literature can be found in [Schmidt, 2005]. This approach complements efforts of the discipline called *voice transformation*, which is the process of transforming the acoustic properties of speech uttered by a source speaker, such that a listener would believe the speech was uttered by a target speaker (as e.g. in [Kain and Macon, 2002]).

Modelling personality exclusively with prosodic features has the advantage that the spoken text is not altered and under full control at design- and runtime. But personality is certainly not only expressed in qualitative attributes of a voice, other properties of a speech dialogue are also essential, like the speaking style and behaviour. For this purpose we have carried out complementary work on an anthropomorphic base grammar that incorporates this aspect, which is described in Section 6.3. This approach guides the development of natural language interfaces with phrasing and behaviour recommendations, which reflect a low or high degree of a personality dimension. Thus, a set of basic phrases can be specialised in several variations, each of which represent a desired personality profile. Within a particular dialogue, the version that suits the personality of an talking object most, can be selected for the actual output generation. A more detailed description of all recommendations and how they were derived can be found in [Hollinger, 2007].

7 EXPERIMENTAL EVALUATION

Towards the goal of fully realising tangible interaction with anthropomorphic smart objects we have encountered new challenges of adequately representing emotions and personality of the corresponding classes. We have explored new ground in the representation of emotions in synthesised anthropomorphic auditory icons and in modelling personality through prosodic parametrisation of synthesised speech [Schmitz et al., 2007b] [Schmidt, 2005]. In the following we will discuss two user studies that explore each of these aforementioned approaches described in Sections 6.1 and 6.2.

In Section 7.1 we outline our efforts to verify whether the created anthropomorphic auditory icons reflect the intended affective states. We further describe in Section 7.2 another user study that has been conducted to find out whether it is possible to model personalities adjusting prosodic parameters of synthetic speech, such that listeners will recognise the intended personality dimension.

7.1 Evaluating the Perception of Anthropomorphic Auditory Icons

In Section 6.1 we have introduced the idea of anthropomorphic auditory icons, which are synthetic affect bursts that can be used for anthropomorphic smart objects. We have described a novel method to construct this type of auditory icons, which basically involves the principle of *vocoding* on existing recordings of human affect bursts as *modulating* sounds and arbitrary synthetic sounds as *carriers*. In this way we are able to use a single custom sound to generate a set of synthetic affect bursts with similar sound characteristics as the original sound sample. In the following we will present findings of a pilot-study exploring the recognition efficiency of emotions in synthetic affect bursts created with this method.

We conducted an explorative online user study with 20 participants of various age groups in order to find out whether the recognition rate of the emotional classes is comparably high as in the studies of the Montreal Affective Voices (MAV), which achieved an average recognition rate of 68%. Furthermore, we wanted to explore whether and how the choice of the carrier sound influences the correct perception of affect.

7.1.1 Method

For this study we created affect bursts created with three different carriers: Sawtooth, sine wave and square wave mixed with sawtooth. The sawtooth-based sounds can be characterised as rattling, wooden, warm while the sine wave is more abstract and mechanic, similar to the robotic sounds of R2D2. The square wave mixed with a sawtooth generator sounds smoother than the pure sawtooth-based samples and has a subtle reverb that makes it slightly more futuristic. *Surprise*, *pleasure* and *sadness* were generated using our own recordings of human affect bursts as modulating sounds, while the remaining categories were synthesised on the basis of the MAVs.

We started off with a pre-study to help us selecting suitable sounds and emotional categories for the actual user study. This rather informal process involved interactive sessions with five persons in each of which we presented the same set of sounds to each participant and noted down first impressions and associations. Based on these statements and the subjective opinions of the authors we have selected six classes of affect bursts: *Anger, disgust, happiness, sadness, surprise, and pleasure*. The former five categories correspond to the Ekman faces, only one Ekman category was excluded from the study (fear), because the pre-study revealed that the corresponding sound samples were difficult to recognise without a given context. Instead, we have included the category *pleasure* to complement our set of synthetic affect bursts, which gives us 18 different sounds in total.

All text was phrased in German and the link to the start page was sent to German native speakers only. At the beginning of the questionnaire participants were asked for gender and age, followed by a page with an embedded Quicktime audio player, which controls a neutral sawtooth-based sound, such that the participant could get used to the handling of the player and test whether the technical requirements are fulfilled. The study consisted of two parts: In the first part one webpage was displayed for each of the 18 sound samples, showing an audio player to control the playback of the sound sample with the possibility to hear each sound as many times as wanted and a text field in which the participants were asked to describe the emotion that they associate with the sound. We explicitly stated that we are not interested in the emotion elicited in the person hearing the sound, but that we want to know what kind of emotion is associated to the entity emitting this sound. The second part was similar except that the text field was replaced with a list showing the six different emotion categories, from which the participant had to choose exactly one. The order of questions within each part was randomised.

7.1.2 Quantitative Results

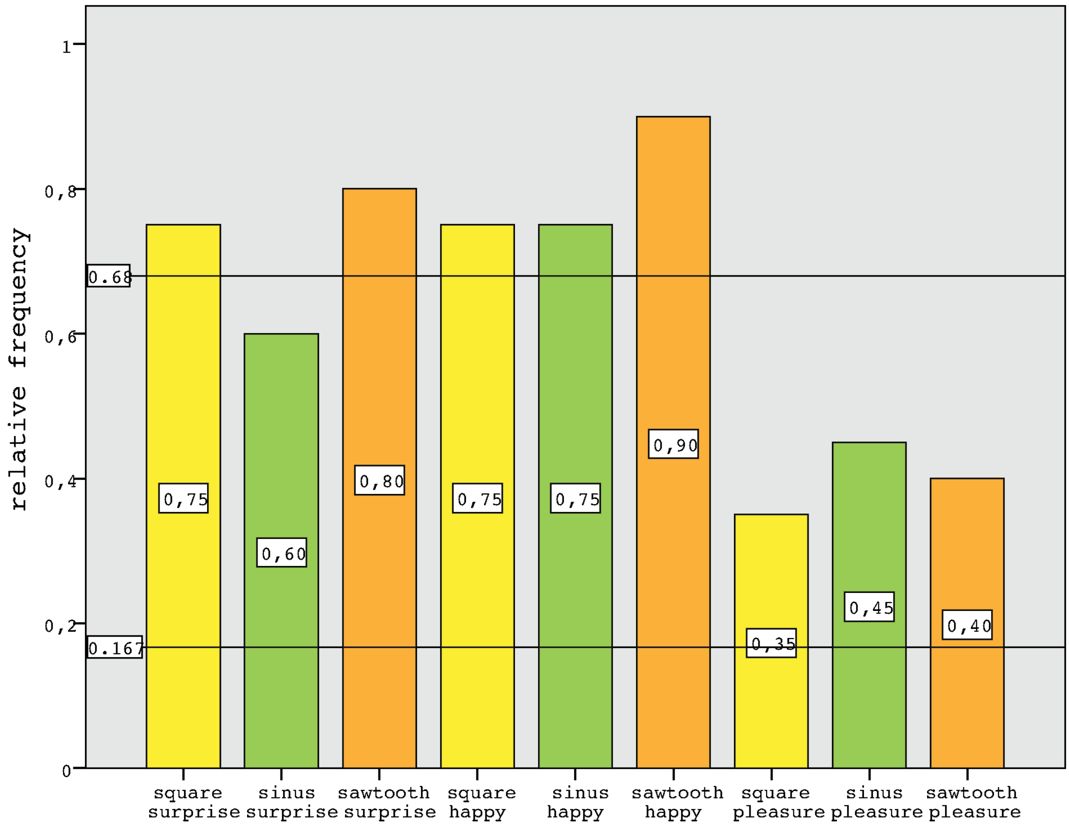


***Figure 7.1**: Mean recognition rates of positive emotions*

A summary of the results of the second part is displayed in Figures 7.1 and 7.2, showing the recognition rates in percent for each synthetic affect burst. The lower horizontal bar indicates the chance level (16.7%) and the upper one represents the average recognition rate of the MAVs (68%). The diagrams shows that out of 18 values only one is below chance level. The sounds representing *surprise* and *happiness* were recognised particularly well throughout all carrier sounds, while the sounds based on the sawtooth carrier obtained the best overall recognition rates. Overall, the positive affect classes (*surprise*, *happiness*, *pleasure*) received better ratings than the negative categories (anger, disgust, sadness).

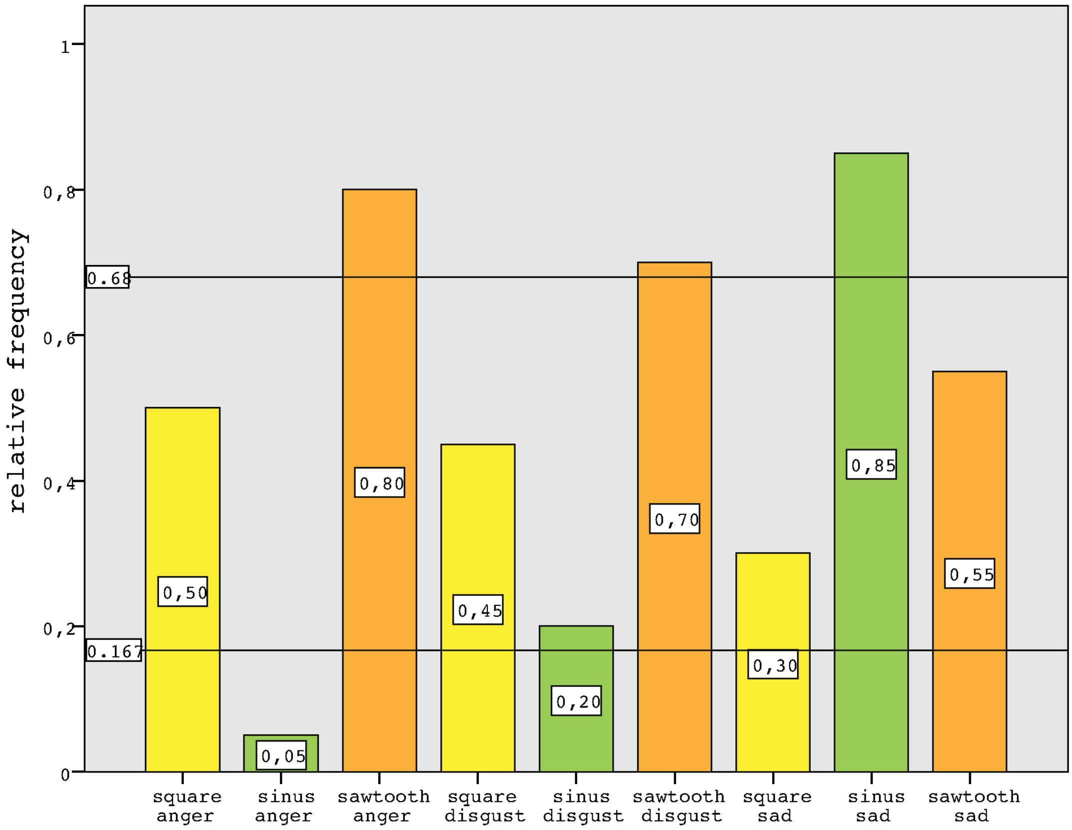


Figure 7.2: *Mean recognition rates of negative emotions*

The category *pleasure* was difficult to recognise among all carriers. The rates of the sine wave carrier are particularly low for *anger* and *disgust*, just as the square wave sound for *sadness*. Other than that, the rates indicate a good overall recognition performance higher than or near to the mean average recognition efficiency of the MAVs.

	Anger	Disg.	Sad	Happy	Pleas.	Surpr.
Square	80	75	65	75	60	85
Sine Wave	55	45	95	85	65	85
Sawtooth	85	100	80	100	65	90

Table 7.1: *Recognition rates for distinguishing only positive from negative emotions.*

Table 7.1 shows the recognition rates when we only distinguish positive (*happiness*, *plea-*

sure, surprise) from negative (*anger, disgust, sadness*) emotions, which means, for example, that selecting *happiness* for the sound that was created to express *pleasure* would be the correct choice.

7.1.3 Qualitative Results

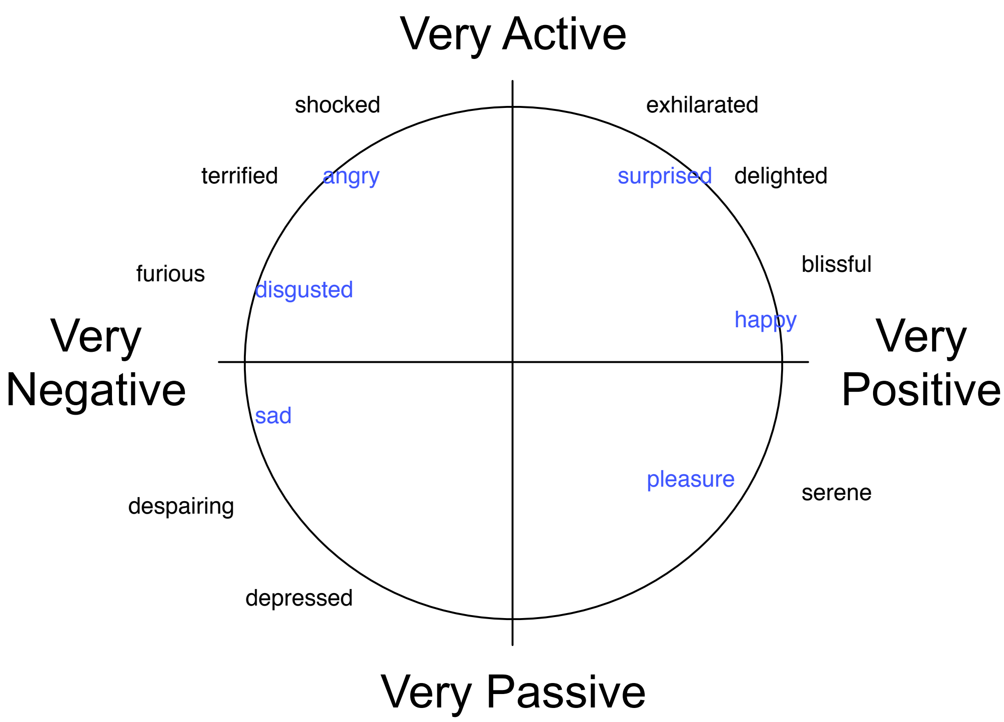


***Figure 7.3**: The activation-evaluation space.*

A first glance at the results of the first part of the study revealed that the majority of the participants had difficulties to phrase terms that clearly describe basic emotions as requested. Instead, participants have freely formulated associations that span from description of events to names of animals. Nevertheless, many of these expressions are related to emotions, such as *"fell down and hurt a knee"* or *"is in a bad mood and stressed"*.

We analysed the results by mapping each statement to a quadrant within the so called *activation-evaluation space*, which is a common representation of emotional dimensions with a long tradition in psychology [Scholsberg, 1941] [Plutchik, 1994]. It represents emotional states by two dimensions, activation and evaluation. Activation denotes how dynamic the state is. For instance, *shocked* involves a very high level of activation, while *depressed* a very low one. Evaluation measures how positive or negative the emotional state is perceived. For example, *happiness* involves a very positive evaluation and *sadness* a very negative one.

Several studies suggest that terms describing emotions can be associated with points in a space based on the these two dimensions. It is further shown that this space is naturally delimited by a circle that is defined by emotional states at the limit of emotional intensity. Thus, each quadrant comprises a cluster of emotions which share the same tendency (positive or negative) in both emotional dimensions activation and evaluation. Figure 7.3 visualises such a space, including a mapping of the emotion categories represented by the sounds in the study.

In order to increase objectivity, two evaluators have checked all statements entered by all participants separately and mapped each of them to one of the activation-evaluation quadrants such that we have two independent ratings of each item. 522 comments were entered in total, including those of users who did not complete the study. In ambiguous situations the statement was omitted from further analyses, which happened with 12.3% of the results for the first evaluator and with 18% for the second one. All ratings were accumulated and summarised in the following tables. The highest rate is displayed in bold font, and the cell representing the quadrant with the intended activation-evaluation tendency is coloured green. Thus, if the green cell contains a bold value, the majority of the participants have recognised the correct tendency in both activation and evaluation.

Sawtooth **Anger**		activity	
		high	low
valence	positive	9.09%	0.00%
	negative	**84.85%**	6.06%

Sawtooth **Disgust**		activity	
		high	low
valence	positive	0.00%	0.00%
	negative	**88.89%**	11.11%

Sawtooth **Sad**		activity	
		high	low
valence	positive	0.00%	9.52%
	negative	28.57%	**61.90%**

Sawtooth **Happy**		activity	
		high	low
valence	positive	**75.61%**	2.44%
	negative	21.95%	0.00%

Sawtooth **Pleasure**		activity	
		high	low
valence	positive	11.11%	37.04%
	negative	**40.74%**	11.11%

Sawtooth **Surprise**		activity	
		high	low
valence	positive	**52.78%**	5.56%
	negative	36.11%	5.56%

Table 7.2*: Activation-evaluation ratings of the sawtooth-based sounds*

Table 7.2 displays results of the sawtooth-based sounds. The intended activation-evaluation tendencies were recognised very well except for the sound representing *pleasure*, which is associated slightly more often to the opposite quadrant. Particularly *anger* and *disgust* stand out with only few wrong associations of the participants.

The set of sounds based on a sine wave achieved ratings as displayed in Table 7.3. The results are rather similar, again all sounds except for *pleasure* were associated most with the intended quality in both activation and evaluation.

Accordingly, Table 7.4 displays results of the sounds that are based on square waves mixed with sawtooth as the carrier. Again, *pleasure* was not recognised as intended, instead the opposite quadrant was chosen most often. Also *sadness* did not achieve higher ratings in the corresponding activation-evaluation quadrant, all four quadrants were generally selected evenly.

Sinus **Anger**		activity	
		high	low
valence	positive	34.48%	0.00%
	negative	**48.28%**	17.24%

Sinus **Disgust**		activity	
		high	low
valence	positive	28.57%	19.05%
	negative	**38.10%**	14.29%

Sinus **Sad**		activity	
		high	low
valence	positive	0.00%	14.71%
	negative	8.82%	**76.47%**

Sinus **Happy**		activity	
		high	low
valence	positive	**56.00%**	4.00%
	negative	24.00%	16.00%

Sinus **Pleasure**		activity	
		high	low
valence	positive	**38.24%**	29.41%
	negative	17.65%	14.71%

Sinus **Surprise**		activity	
		high	low
valence	positive	**60.00%**	13.33%
	negative	26.67%	0.00%

Table 7.3*: Activation-evaluation ratings of the sounds based on sine waves*

Square **Anger**		activity	
		high	low
valence	positive	12.90%	0.00%
	negative	**67.74%**	19.35%

Square **Disgust**		activity	
		high	low
valence	positive	41.67%	0.00%
	negative	**54.17%**	4.17%

Square **Sad**		activity	
		high	low
valence	positive	25.71%	**28.57%**
	negative	20.00%	25.71%

Square **Happy**		activity	
		high	low
valence	positive	**61.76%**	0.00%
	negative	38.24%	0.00%

Square **Pleasure**		activity	
		high	low
valence	positive	17.65%	17.65%
	negative	**52.94%**	11.76%

Square **Surprise**		activity	
		high	low
valence	positive	**65.52%**	6.90%
	negative	24.14%	3.45%

Table 7.4*: Activation-evaluation ratings of the sounds based on square waves mixed with sawtooth*

7.1.4 Discussion and Outlook

Looking at the quantitative part, the sine wave seems to be a limited carrier for certain emotions (*anger* and *disgust*) in direct comparison to the other carriers. An explanation might be that it only carries energy in limited frequency bands, such that only insufficient acoustic properties for these particular emotions could be transferred to the carrier sound. The mere pitch progression and intensity curve is not sufficient to recognise these synthetic affect bursts. The distinct "growl" of the recorded sound for anger is only recreated in carriers that cover a wide frequency spectrum, which is assumed to behave similar with the emotion *disgust*. *Sadness* works best with the sine wave carrier, for which we suspect that the pitch curve (slowly decreasing) is most important. The square wave slightly diffuses the pitch curve due to its reverb. The recognition of *Surprise* seems to rely on pitch and intensity (both raise quickly and then decrease slowly). *Happiness* is particularly recognisable since the typical rhythm of laughter can be recognised in all carriers equally well. *Pleasure* is very similar to *surprise* in its pitch and intensity curve and semantically very close to *happiness*, which would explain the comparably low recognition rates.

This online study (intentionally) did not provide any context for the sound, which is certainly different in real world use cases. In everyday encounters, the context of the interaction is an important factor when potentially ambiguous affect bursts have to be interpreted.

Furthermore, concrete applications are usually set in a particular domain and interaction context and thus provide clues that help to interpret acoustic feedback more fine grained and to resolve ambiguities. For instance, a sound expressing *disgust* will probably be less often interpreted as *sadness* in scenarios that involve food or beverages compared to context-free situations.

Further, if we group the six categories into a positive and a negative emotional region, it shows that the distinction between good and bad is very accurate throughout most stimuli. This is an essential feature of such synthetic sounds, since in real world applications differentiation between positive, encouraging feedback versus negative, refusing ones is crucial and typically more important than resolving nuances between e.g. anger and sadness.

The qualitative part of the study did not even present any categories to chose from, obviously increasing the degree of difficulty to interpret the sounds. Nevertheless, reducing the discrete emotional categories to four activation-evaluation quadrants has shown that participants were generally able to correctly associate emotional value that is at least "nearby" the intended one. One exception in this part of the analysis is the sound representing *pleasure*. One reason might be that this emotion is commonly expressed with lower intensity than the remaining ones, and therefore more prone to variations in nuances when perceived in real life. Repeating the same procedure of building emotional clusters according to activation-evaluation quadrants with the data we got in the second part of the study - when all six emotional categories were given in a drop-down list - we can see that almost all sounds were easier to rate for the participants. Out of 18 sounds 14 show an increase in their recognition rate when emotional categories are given, 1 did not improve and 3 became slightly worse.

This experimental user study clearly indicates that our novel concept of anthropomorphic auditory icons is able to produce sounds that can express emotions effectively. As a first cornerstone it paves the ground for future efforts to further improve this approach. Problematic emotional categories have to be revisited and additional ones could be included. It might further be interesting to isolate acoustical parameters such as the pitch curve and to particularly observe effects of variations on the perception of affective content.

Another major step in the development process of anthropomorphic auditory icons will be the possibility to further parameterise such sound artefacts, which would allow to encode additional information. To complete the compilation of life-like sounds physiological sounds could be considered as well, representing organic states or activities, such as sounds of swallowing, groaning, clearing one's throat, or heart beat. Such sounds could also be employed as idle-time actions in order to maintain the anthropomorphic stance (see also Section 4.1.4).

7.2 Perception of Personality in Voices of Anthropomorphic Products

To verify our assumptions about prosodic parametrisation of speech output for talking objects with personality described in Section 6.2, we conducted a user study in order to evaluate whether it is possible to model different personalities with the same voice by adjusting prosodic parameters, such that listeners will recognise the intended personality dimension. We also intended to determine, whether a product itself is associated with a personality just by seeing it.

7.2.1 Method

For this study 12 audio files were generated in total (for each gender 5 personalities plus baseline), saying ”Hello, I am product XY. I would like to introduce myself. I will now explain my features.” (English translation) and all played in random order to each participant. After listening to a stimulus, the participant had to rate the voice on a Likert scale from 1 (does not fit at all) to 5 (fits very well) in each personality dimension

In the second part of the study a variety of pictures were presented one by one to the same group of participants, each showing one product (bottle of wine, hammer, DVD player etc.) that we expect to represent one brand personality dimension. As with the voices, participants had to express the perceived object personality by rating each picture (see Figure 7.4) on all brand personality dimensions on a Likert scale from 1 to 5.

***Figure 7.4**: Product pictures of the second part of the study*

Care was taken to ensure that products are chosen from a large variety of categories, in order to be able to derive general conclusions as much as possible. Judgements were done as in the first part of the study, complemented by comment fields for free remarks. The sequence of voices and products was varied in order to avoid sequence effects in the results, which are basically interferences due to a fixed order of questionnaire items [Hager, 1987].

Each run required approximately 30-40 minutes. The procedure was explained to the participants and statistical data such as age, gender and level of experience with computers was collected prior to the actual study. A short comment generated with the speech synthesiser was presented such that participants get used to the artificial voice. It was also stated that it is not relevant *what* the voice is speaking but *how* it sounds. 36 people participated with the study, of which 25 were between 17 and 27 years old, 7 between 27 and 37 years and 4 persons were older than 37. 26 participants were male and 10 female.

7.2.2 Results

For all dimensions the means of all ratings of the voices were taken and then compared on each dimension. The analysis of the ratings of voices shows that for all personality dimensions differences there exist highly significant differences between at least two voices, with the p-level of the ANOVA being below 0.001 for all dimensions. The Fisher-LSD-Test provides a more detailed overview on the voices that show significant results, of which we will summarise the most important aspects in the following paragraphs. Besides that we conducted an ANCOVA (Analysis of Covariance) with gender as the additional covariant, which means that we have analysed the ratings of male and female participants separately as well. But, since the differences of this ANCOVA were negligible except for the dimension *competence*, we will only consider gender specific results in the section related to the rating of *competence*.

The results of the voice evaluation are displayed in Figures 7.5 to 7.10. The scale is set from 1 (”not at all”) to 5 (”very much”) with a midpoint of 3 (”neutral”), which is marked in the figures. The *Baseline* specifies the neutral voice. To measure if a voice has been successfully recognised, we have considered the average rating of a voice with respect to the neutral value of 3, to the baseline and to other voice ratings.

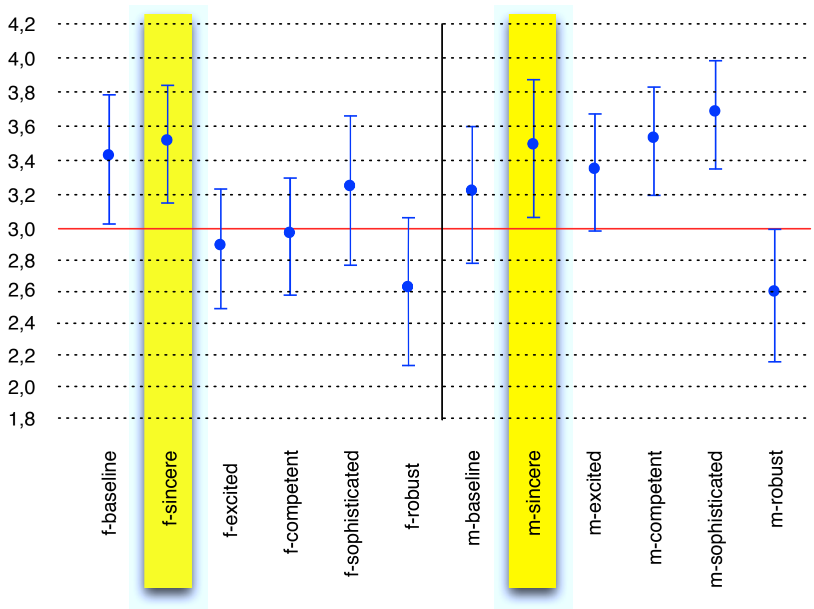


Figure 7.5: *Mean averages for sincerity, vertical bars show confidence intervals (0,95)*

Sincerity: As depicted in Figure 7.5, sincerity is generally well recognised in the respective voices. The voices modelled with high sincerity achieved a rating that is higher than the neutral voices and also clearly higher than the mean value. Among the female voices, the sincere one is the voice with the highest average rating in this dimension. Among the male voices, the sincere voice is also rated as comparably sincere, although the male sophisticated voice obtained an even higher one.

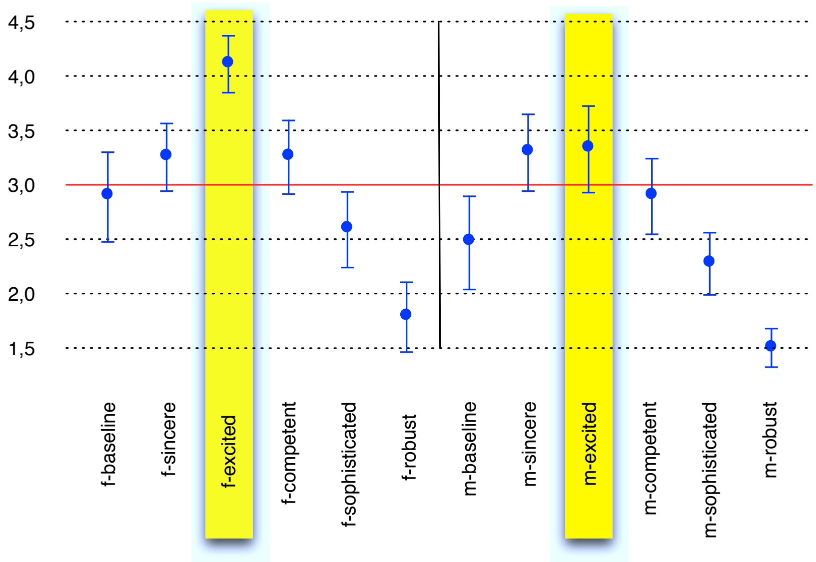


***Figure 7.6**: Mean averages for excitement, vertical bars show confidence intervals (0,95)*

Excitement: Figure 7.6 shows that excitement is generally very well recognised in the 2 corresponding voices. The rating of the excited voice is far above the mean value and higher than the neutral voice. Particularly the female version of the excited voice is clearly recognised as such. Also the male voice is the one that is rated as the voice with the highest value of excitement. The rugged voices have the lowest rating in this dimension. The differences between rugged and excited voices are primarily tempo and pitch level, which supports the hypothesis that increased pitch level (+1) and tempo (+1) can reflect excitement.

Competence: Figure 7.7 indicates that the ratings for competence are generally higher for male voices than for female voices. The male competent voice has been recognised well - its rating in competence is higher than the mean value and higher than the neutral voice. But, the highest value in competence is achieved by the male cultivated voice and the highest rated female voice is the one modelled for sincereness. The female voices for the lowest rating in competence are the ones created for ruggedness and excitement. The male voices for sophistication and competence have more in common than the gender, including a pitch level of normal and low, a high pitch range and a tempo range from normal to slightly faster.

The differences between the ratings of male and female participants are apparent in a direct comparison as in Figure 7.8. The female competent voice is rated much lower than the male competent voice by male participants. Thus, female voices are not generally perceived as less competent but only when judged by male persons. An obvious consequence would be to always chose a male voice if competence should be expressed and if the target group includes male people.

Sophistication: As depicted in Figure 7.9, sophistication is particularly well recognised

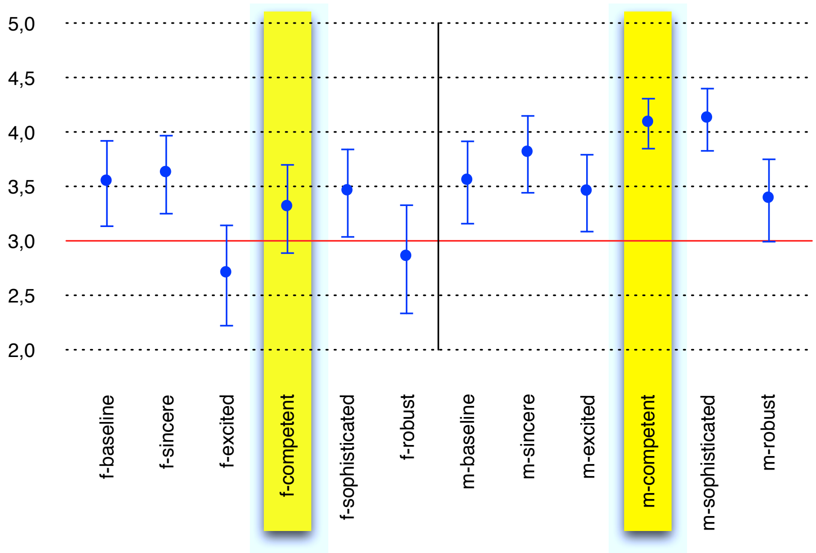


Figure 7.7: *Mean averages for competence, vertical bars show confidence intervals (0,95)*

in male voices. Both male and female versions are rated higher in sophistication than the baseline and are above the mean average. Furthermore, the male sophisticated voice obtained the highest rating in sophistication among all voices. The rugged voices (female and male) had the lowest ratings in sophistication. The sincere voices achieved a rather high rating in sophistication as well, particularly the female sincere voice got the highest rating. Common attributes of the voices modelled as sincere and sophisticated are an increased pitch range and a normal tempo, which are also the two parameters that differ most from those of the rugged voice.

Ruggedness: The dimension of ruggedness also reveals clear differences between female and male voice as seen in Figure 7.10. Female voices are generally rated lower in ruggedness than the male counterparts. Accordingly, the female rugged voice is not perceived as rugged. Thus, the gender seems to play a major role in the perception of this dimension as well. The male version has been rated higher in ruggedness than the neutral one with a rating above mean average. But the male competent voice achieved a clearly higher rating than the actual rugged one. The excited voices are the ones rated lowest in ruggedness, which further supports the above assumption that the most relevant attributes to reflect excitement and ruggedness are pitch level and tempo. Both male and female rugged voices exhibit most significant differences to all other voices. Each of them are significantly different to 8 out of 11 voices.

Correlation Analysis: With the correlation analysis of the study data we attempt to derive dependencies of voice attributes and the perception of perceived personality dimensions. The correlation values are illustrated in Table 7.2.2. Gender is coded numerically, such that male is denoted by 0 and female by 1. Clearly significant correlations are marked red in the table and summarised in the following:

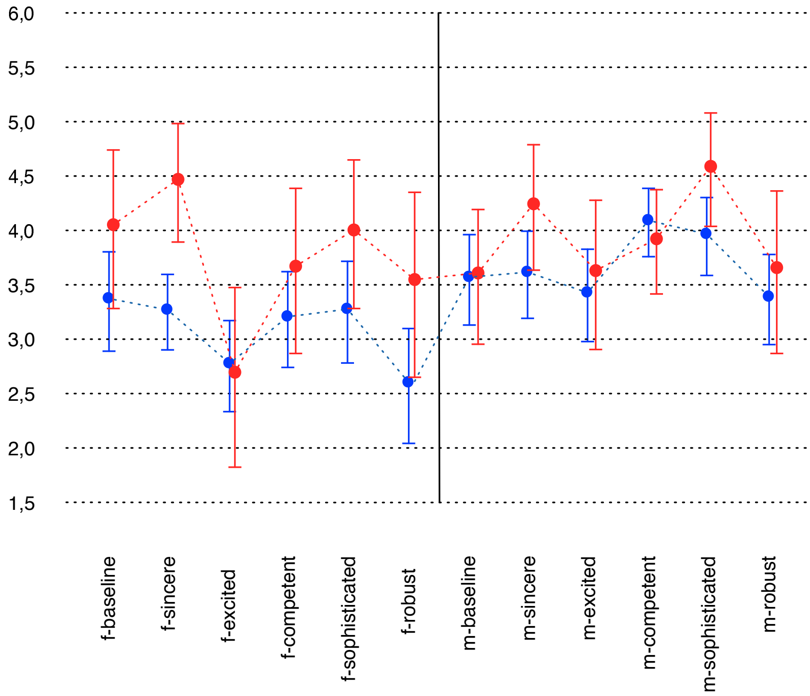


Figure 7.8: *Mean averages for competence by gender (red = female, blue = male), vertical bars show confidence intervals (0,95)*

- Excitement correlates with pitch level, pitch range and speed. Thus, increasing these values will yield an increased perception of excitement.
- Competence correlates with gender. Male voices are generally perceived as more competent.
- Sophistication correlates with pitch range. Voices with higher pitch range will be perceived as more sophisticated.
- Ruggedness correlates with pitch and gender. Lower pitch is perceived as more rugged and male voices are perceived as more rugged than female.

7.2.3 Summary and Discussion

Tables 7.6 and 7.7 show how the parameterised utterances were rated in each personality dimension (arithmetic mean). It can be observed that in most cases the adjustments increased the rating of the voice in the intended dimension in comparison to the baseline, while the male version worked better than the female one. The best rated versions of the utterance in each dimension (bold values) are not always the intended ones, e.g. the male competent voice received the highest score in the dimension rugged.

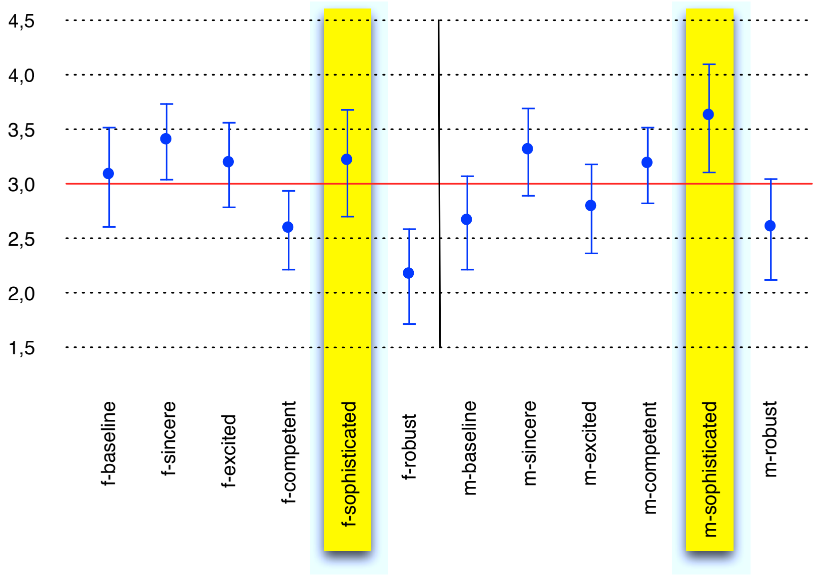


***Figure 7.9**: Mean averages for sophistication, vertical bars show confidence intervals (0,95)*

	Correlations (p<0,05)				
Variable	Pitch	Range	Rate	Volume	Gender
Sincerity	0.14	0.52	0.41	-0.35	-0.29
Excitement	0.73	0.62	0.79	0.25	0.24
Competence	-0.26	0.27	0.12	-0.31	-0.64
Sophistication	0.23	0.63	0.38	-0.20	-0.07
Robustness	-0.62	0.11	-0.05	-0.09	-0.69

***Table 7.5**: Correlation analysis*

The ratings of the products confirmed our assumptions clearly, basically all products got the best rating in their intended dimensions. The ratings were consistent among most participants and are significant with almost all products.

Both studies have shown that there are clear preferences for our prosody-modelled speech synthesis for certain brand personality dimensions. Interdependencies between certain personality dimensions are probably also a reason why it is difficult to adjust prosodic parameters in order to reflect one dimension only: Excited and rugged are clearly contrary dimensions already by definition, reflecting in the results as well, since voices with a high rating in one of the two dimensions get a low score in the other one. In some cases, the intended dimension did not get the highest rating, but compared to the higher rated voice there are still characteristic distinctions if the values of the remaining dimensions are considered as well: For example, the rugged voice was topped by the sophisticated one in the dimension rugged, but looking at the last two rows of Table 7.6 you will see that the other values of the

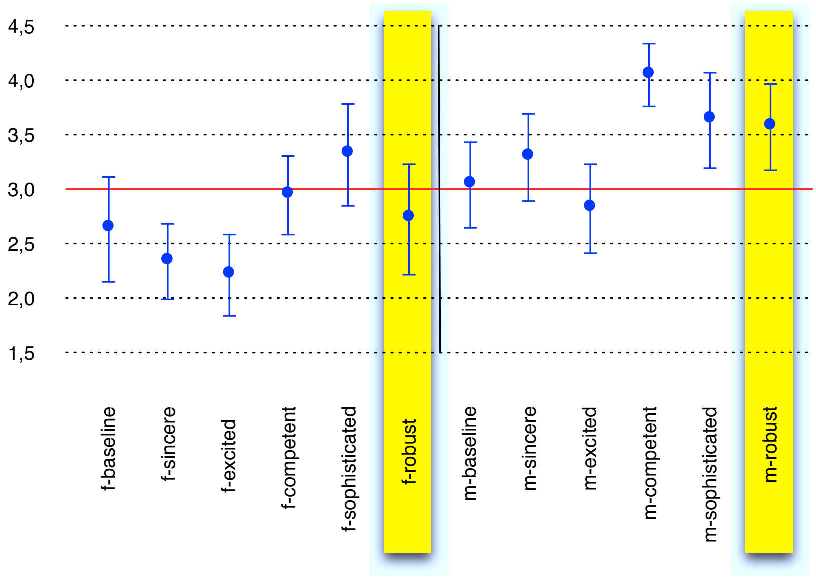


***Figure 7.10**: Mean averages for ruggedness, vertical bars show confidence intervals (0,95)*

assigned to / modelled as	sincere	excited	compet.	sophist.	rugged
Baseline	3.2	2.5	3.5	2.7	3.0
Sincere	**3.5**	3.3	3.8	3.3	3.3
Excited	3.3	**3.4**	3.5	2.8	2.8
Competent	3.5	2.9	4.0	3.2	**4.1**
Sophist.	**3.7**	2.3	**4.1**	**3.6**	3.6
Rugged	2.6	1.5	3.4	2.6	**3.6**

***Table 7.6**: Mean values of the male voice*

rugged voice are much lower than the ones of the sophisticated voice, such that in the rugged voice the rugged dimension is relatively prominent, which is not the case with the sophisticated voice, where only one other dimension is rated lower than the rugged one. Therefore the rugged voice still generates a suitable, rugged personality profile. The gender of the voice also has a strong effect on the perceived personality, female voices generally get lower values in competency and ruggedness and score higher in excitement, it would therefore be suggestive to chose the gender depending on the personality you would like to express.

We also learned from this study that there is still room for improvement in the prosodic variation of voices in order to express certain personality features:

- Gender appears to be vital, certain personality dimensions could not be expressed with voices of both genders. Also, in future research prosodic parametrisation should be

assigned to / modelled as	sincere	excited	compet.	sophist.	rugged
Baseline	3.4	2.9	3.5	3.1	2.6
Sincere	**3.5**	3.2	**3.6**	**3.4**	2.4
Excited	2.8	**4.1**	2.7	3.2	2.2
Competent	2.9	3.2	3.3	2.6	3.0
Sophist.	3.2	2.5	3.5	3.2	**3.3**
Rugged	2.6	1.8	2.8	2.2	2.7

***Table 7.7**: Mean values of the female voice*

realised in a gender specific manner and therefore regard female and male voices independently.

- The female sincere voice was well recognised and therefore does not need to be adjusted.

- The male sophisticated voice also received the highest rating in both dimensions sincerity and competence. The major difference between the sophisticated and sincere/-competent voices is loudness, which is higher for sincere and competent voices. The male voice with the quality of normal loudness seems to be preferred over the one with higher loudness. Since the speaker was requested to increase the volume when recording for the diphone inventory, not only the volume increased by means of higher decibel but also the general voice quality, which might interfere with other prosodic parameters and constitute a potential cause for this result. Thus, we suggest to adjust the male sincere and competent voices to normal loudness.

- The excited voices were recognised very well and do not require changes.

- Since the sophisticated male voice received a higher rating in competency than the competent one, it should be adjusted as outlined above. Female voices did not successfully reflect competence, particularly the voice modelled to reflect competence had a lower rating than other female voices with a higher pitch, which suggests to increase the pitch level from -1 to 0.

- The male sophisticated voice was recognised well and does not need to be adjusted. The female counterpart could not achieve the highest rating, which is possibly caused by a low pitch level, since the sincere voice with normal pitch has obtained the highest rating in sincerity. Therefore, increasing the pitch level of the female sincere voice from -1 to 0 might be appropriate.

- In order to express ruggedness, male voices are more adequate, since female voices were perceived as less rugged overall. Furthermore, the male rugged voice seemed to be too slow. Participants' comments related to this voice have revealed that voices appear to be dislikable when they are slow. Voices that are faster but also low pitched, such as the sophisticated and competent male voices, are perceived as equally rugged and even more rugged. Thus, it is recommended to increase the tempo of the rugged voice from -1 to 0. The voices that are originally intended to be modelled as the rugged ones have the property that they received a particularly low rating in all other dimensions except the rugged one.

Applying these lessons learned from the study, we obtain a new parametrisation for female and male voices respectively, which is illustrated in Tables 7.8 and 7.9.

Personality Dimension	Pitch level	Pitch range	Tempo	Loudness
Baseline	0	4 st	+15%	Modal
Sincere	0	8 st	+15%	Loud
Excited	+30%	8 st	+30%	Loud
competent	0%	8 st	+30%	Loud
sophisticated	0%	8 st	+15%	Modal
rugged	-	-	-	-

Table 7.8: *New parameter matrix for female voices, changes are red*

Personality Dimension	Pitch level	Pitch range	Tempo	Loudness
Baseline	0	4 st	+15%	Modal
Sincere	0	8 st	+15%	Modal
Excited	+30%	8 st	+30%	Loud
Competent	-30%	8 st	+30%	Modal
Sophisticated	-30%	8 st	+15%	Modal
Rugged	-30%	4 st	+0%	Loud

Table 7.9: *New parameter matrix for male voices, changes are red*

Other voice parameters can be applied in future to even increase the perception of personality dimensions, such as modelling age. This could for instance be used to increase the impression of excitement by applying a younger voice and sophistication through an older one. Speech qualities such as huskiness might deliver a sense of ruggedness. Accents could also open new possibilities to underline attributes related to personality, for example by employing stereotypes which fit to the object, such as French or Italian accents for the corresponding types of wine.

To our knowledge this is the first attempt to model brand personalities by mapping prosodic parameters of synthetic speech to personality dimensions, which already works quite well for some cases. Since associating brand personalities to products has proved to be much easier in the second part of the study, we expect that the combination of both (seeing a product that talks with a matched voice) will increase the overall effect and provide clearer results. More details about this user study can be found in [Schmidt, 2005].

7.3 Synopsis

The explorative studies presented in this chapter successfully demonstrated the efficiency of our novel approaches for the expression of personality and emotion in different categories of TASOs.

In an initial experiment we have studied the effects of three sets of anthropomorphic auditory icons that we have created exemplarily. While not all emotion categories seem to be represented sufficiently, the majority of the tested sounds performed very well in this study. The results clearly encourage to pursue this approach, constituting the first conceptualised methodology of creating such synthetic affect bursts.

The findings of the user study on parameterised speech generally provide indications of how to convey certain personality traits with diphone synthesis. This is the first attempt to model brand personalities by mapping prosodic parameters of synthetic speech to personality dimensions, which already performs well for some cases. Modelling personality exclusively with prosodic features has the advantage that the spoken text is not altered and under full control at design- and runtime and that a sense of personality can already be conveyed in short interaction sequences.

8 APPLICATIONS

We have introduced the concept of TASOs in Chapter 4 and proposed several complementing types of anthropomorphism that could be applied in various scenarios. In addition to that we have developed novel methodologies for generating vocal behaviour of anthropomorphic objects, which can be employed as building blocks for creating instances of TASOs. Chapter 5 described our prototyping architecture for realising TASOs as part of AmI applications, which forms the technological foundation of the application prototypes discussed in this chapter. Although not all aspects of the proposed concepts and technologies could be reflected in all prototypes, the developed systems demonstrate the manifold potentials.

We will start with describing the *Sensing Pill Box*, which leverages sensor integration for the creation of digital product memories, supporting patients' compliance regarding the intake and storage of medicine. We will continue with the *Digital Sommelier*, a shopping application that involves tangible interaction with anthropomorphic bottles in an instrumented shelf. Finally, we introduce the *Anthropomorphic Cocktail Shaker*, an interactive shaker that guides through a cocktail shaking process, setup as an edutainment installation.

8.1 The Sensing Pill Box

Emerging sensor technologies that go beyond RFID recognition are becoming cheaper and smaller, which will also significantly affect future generations of products as elements of production, processing and delivery chains, enabling them to sense and record their state and communicate with their environment. The prospective immediate benefits for manufacture and retail are an essential factor that is increasing the chances for research work in this field to get adopted by industry - just as RFID technology has already made the step into the commercial sector, as it provides economic values by facilitating process automatisation such as product tracking or inventory stocktaking.

The SemProM project (as introduced in Section 2.1.3) focuses on utilising product histories in order to facilitate value-added services in different stages of a product's life-cycle. These product histories are also called *Digital Product Memories* (DPMs) and enable novel applications along the supply chain and the product's life cycle, respectively (see e.g. [Kröner et al., 2009]). The basic concept of SemProM aims at new strategies for the "Internet of Things". It is based on semantic technologies, machine-to-machine-communication, intelligent sensor networks, instrumented environments, radio-frequency identification technology and multi-modal interaction.

Within the SemProM project we have realised the Sensing Pill Box. Figure 8.1 on the right shows how we have augmented an ordinary pill box by simple means with a μPart (see Table 5.1.1) into the cap. The sensor module is mainly integrated for two purposes: First, environmental conditions (i.e. the current temperature) are observed and verified whether they are within the specified limits. Second, the intakes of the pills by the patient are recorded and logged. This is realised by checking the light sensor of the integrated sensor module, which recognises changes of the light level when it is opened. Although this sensor module is already uniquely identifiable, we additionally tagged the object with a RFID chip.

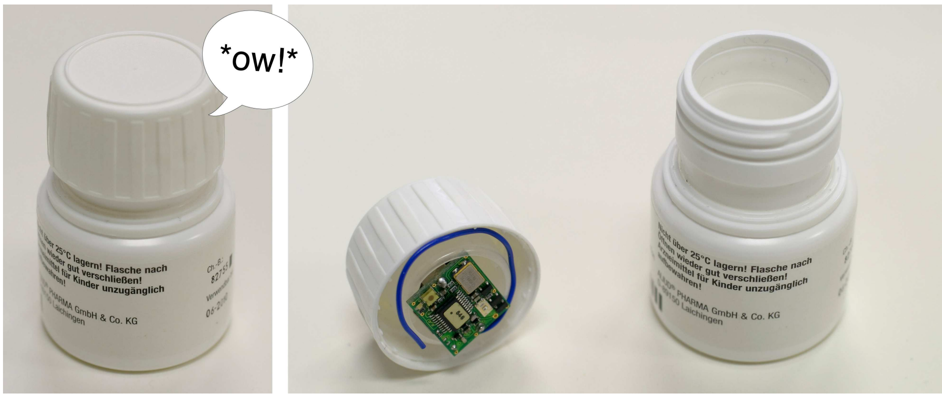


Figure 8.1: *The Sensing Pill Box complaining when exposed to heat (left), the integrated sensor (right)*

Such an augmented pill container that is uniquely identifiable already enables additional functionality much earlier within its life-cycle - for example during the production phase conducted in an instrumented production plant, such as the SmartFactory$^{\mathrm{KL}}$ (e.g. [Zuehlke, 2010]). The SmartFactory$^{\mathrm{KL}}$ has the purpose of supporting the development, application and propagation of innovative industrial plant technologies in different economic branches as well as providing a basis for their extensive usage in science and in practice. Projects realised in this living lab include machines that sense and identify containers at certain steps of a filling process (see Figure 8.2 at the right). The filling configuration of the pill box is stored at the item itself, thus the pill box identifies itself at the filling machine and demands from the machine to be filled with a particular setup of pills. These information and also the actual filling events are stored in the object's DPM and are available for retrospective processes. Although such behaviour occurs in machine-to-machine interaction, it already employs an anthropomorphic pattern on the system design level.

Figure 8.2: *The pill box memory visualised by the SemProM browser (left). Filling process in the SmartFactory (right).*

At home, the sensory equipment enables further novel services. The Sensing Pill Box application software typically resides on a common desktop PC and receives the sensor data, which will be abstracted and stored into the DPM of the pill box. The DPM can later be visualised by other software components, such as the SemProM-browser [Kröner et al., 2009], see Figure 8.2 for a picture of a presentation of the SemProM-browser showing the DPM of a pill box at the CeBIT fair in 2009). This abstracted data consists of a higher level interpretation of the events, which for instance describe that the pill box has been stored under inappropriate conditions or that the actual intake of the medicament by the patient has been delayed by a certain amount of time. This information can be used to support pro-active assistance systems that e.g. remind the patient to take the prescribed dosage as soon as possible. Such a product memory could also inform the doctor about the patient's compliance.

For an end-user scenario at home we have assembled affect bursts to reflect certain states of the pill box. Rendering these sounds over loudspeaker systems at home informs the user

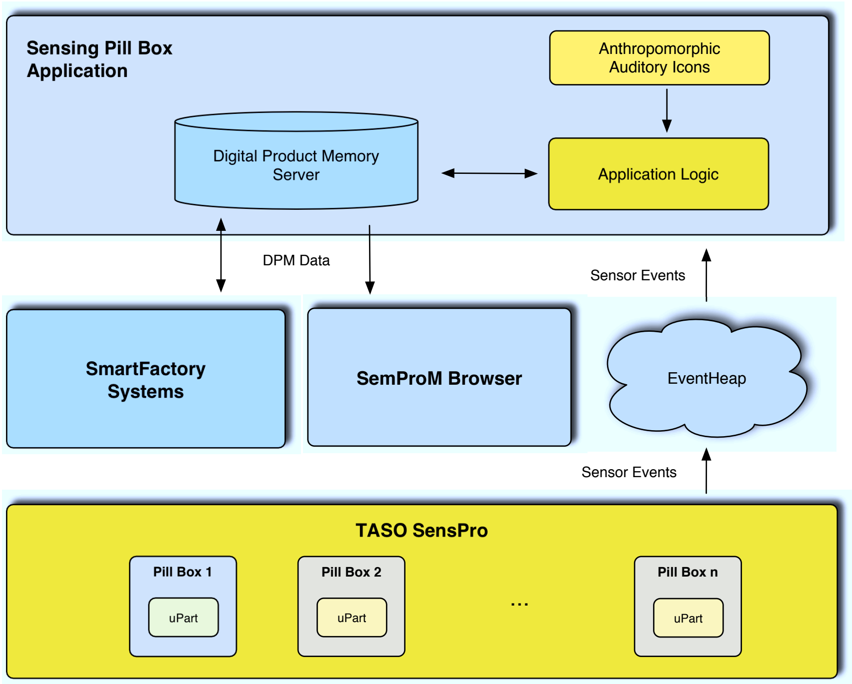


Figure 8.3: *Architecture of the Sensing Pill Box based on the TASO Platform.*

about incidents that have to be attended or acknowledges that certain user actions have been recognised. Figure 8.3 provides an overview of the architectural setup of the Sensing Pill Box. In the following list we describe all events that are identified by the Sensing Pill Box and how they are mapped to affective non-verbal output of an animalistic smart object (as defined in Section 4.3.1).

Pain: This indicates that the medical container is stored under inappropriate conditions (Temperature is not within predefined limits, or medical contents are exposed to light, because the container has not been closed after last intake). See also Figure 8.1 on the left.

Happy: The movement sensor detects when the pill box is picked up. Although this interaction event does not yet reflect user actions relevant to compliance, it confirms that the Sensing Pill Box is active and online.

Pleased: Whenever the pill container is opened and the next intake as prescribed by the doctor is pending, this type of positive feedback aims at encouraging the patient's activity, assuming that the box has been opened to retrieve a pill.

Fear: As opposed to the previous state, it might also be the case that the patient opens the container at the wrong time, which will result in this warning. The patient is alarmed and can decide on further steps.

Swallowing: When the pill box has been closed, the system of the Sensing Pill Box logs this activity as an intake at that particular time. Acknowledging this logging action to the user is represented by a sound of swallowing.

Since the intake of medicine is a reoccurring task, such affect bursts / anthropomorphic auditory icons will provide a relatively quick and less obtrusive feedback on various situations, compared to - for instance - extensive speech output. If the user is not clear about the state of the system, further means to inspect the DPM should be provided, such as described in [Brandherm et al., 2010], which realises authorised and personalised access to DPMs.

Summing up, the Sensing Pill Box realises anthropomorphism in smart objects by employing a semantic memory, enabling self-configuration by articulating demands to other machines, and through animalistic pro-active system responses and reactions to the patient's behaviour at home.

8.2 The Digital Sommelier

Embedding sensors into products to enable Digital Product Memories as outlined in the previous section is also the foundation of the application presented in the following. As we have seen, sensor and communication technology that continuously sense and observe the state of a product open up possibilities for new services that also benefit end consumers: Another simple example is quality assurance of sensitive food such as fresh fish, by monitoring the continuous maintenance of the cooling chain. By providing information about the temperature history to end customers, their trust into the product could be increased. Temperature is only one product attribute that could provide useful information to the customer, we envision that DPMs will be able to record a complete digital diary with all relevant information, particularly regarding their quality and operation. Being able to communicate with their environment it is further important to allow humans to access such a product memory.

***Figure 8.4**: The DigiSom installation in Shenyang, China*

We developed a demonstration application, the *Digital Sommelier* (DigiSom) [Schmitz et al., 2008a], that provides a simple interface for customers to retrieve general features of the product (several types of wine and champaign) as well as specific state attributes: The current temperature and whether the bottle has been recently shaken. One goal was to support novice users in their shopping activities with an intuitive interface. The system recognises user interactions based on RFID and sensor technology, and provides natural language combined with visual output of information. We also integrated our concept of talking products with personality (see Section 4.3.2) into this scenario: Each type of wine is associated with a distinct personality to represent a sense of brand and wine type. Personalities are reflected by prosodic attributes of the voice (see Section 6.2) and by different phrasing styles (see Section 6.3 and Appendix A.2). The implementation of the sensor integration is based on TASO SensPro as described in Section 5.1. The DigiSom installation has been exhibited on various occasions (e.g. at the exhibition series called ”Deutschland und China - Gemeinsam in Bewegung” organized by the Federal Foreign Office of Germany presented in China - see Figure 8.4) and is now a permanent installation at the Innovative Retail Lab.

Figure 8.5: *Talking bottles that greet and introduce themselves*

8.2.1 The DigiSom Installation

In a shop the user has the possibility to communicate with the offered products, as her actions in the store are recognised and monitored by a combination of RFID and sensor technologies. The shelves in the store are equipped with RFID-readers, and every product is tagged with passive RFID-transponders, which allows for locating and identifying products. Every RFID-transponder also contains a reference to all relevant product information, including static information such as its name, ingredients and nutrition facts and also dynamic data like a temperature history that accumulates over the product’s lifetime.

In addition to that the DigiSom employs various sensors either directly attached to the wine bottles or in case of champagne embedded in the product box, in order to gather information about the state of a product, i.e. the actual temperature of the bottle of wine and whether the bottle has been shaken or not. Furthermore, interactions with the product box can be monitored using the attached sensors, which are able to measure light, temperature, sound, and acceleration.

Whenever the user takes a bottle out of the shelf the attached RFID-tag disappears from the RFID-reader field and the DigiSom application receives an event from the EventHeap, which triggers the generation of the according product information page to be displayed on a nearby screen assigned by the Presentation Manager and at the same time initiates the

Digitaler Sommelier

Produktinformation

Produkt:	Gosset Grande Réserve Brut
Jahrgang:	Cuvée verschiedener Jahrgänge
Alkohol:	12,00 Vol%
Preis:	30,10 EUR
Hersteller:	Gosset
Website:	www.champagne-gosset.com
Herkunftsland:	Frankreich, Champagne
Traube:	Chardonnay (46%), Pinot Noir (39 %), Pinot Meunier (15 %) und "Reserve-Weine"
Serviertemperatur:	7 °Celsius
Farbe:	Dunkles Gold-Gelb, leichte kleine Perlen
Beschreibung:	Intensiv und komplexe, aromatische Note Blumig (Lindenblüte, Narzisse) und fruchtig (Kirsche und Brombeere) Am Anfang Aroma von getrockneten Feigen und glasierten Kirschen dann geröstete Mandeln und Zwieback Langer, kräftiger Abgang
Passt zu:	Geeignet als Aperitif Champagnercreme

Tipp:

Die aktuelle Temperatur des Produktes beträgt: 23 °Celsius
Sie sollten das Produkt vor dem Servieren kühlen!

***Figure 8.6**: The generated product page of a Champagne bottle*

champagne to introduce itself using a voice with a prosodic parameter set expressing a particular brand personality (see Figure 8.5). The current implementation of the DigiSom renders all speech through SAFIR (see Section 5.3.3), which theoretically allows to position the virtual sound source depending on the product's position, if it is available to the system. Furthermore, the application receives the current temperature of the champagne from the integrated sensor node via TASO SensPro, comparing the actual temperature against the proposed temperature at which the champagne should be served. When the current temperature is not within the recommended temperature range, it generates appropriate hints, e.g. that the champagne should to be cooled down before serving (see Figure 8.6).

If the user turns the champagne to read the information on the rear side of the box, the associated sensor detects the rotation and generates the associated event, which is passed to the EventHeap. The DigiSom interprets this action as an indication of the user's interest in the product, such that the DigiSom changes the information currently being displayed. In this case the application refers the user to the web page of the manufacturer, where the user can browser for additional information about the product (see Figure 8.7).

Through the combination of information from RFID-tags, data about the champagne's/wine's actual temperature, and whether the bottles have been shaken or not, as well as acceleration information from Crossbow motes (see Table 5.1.1), the DigiSom is able to incorporate real time data about the state of products to adapt product presentations.

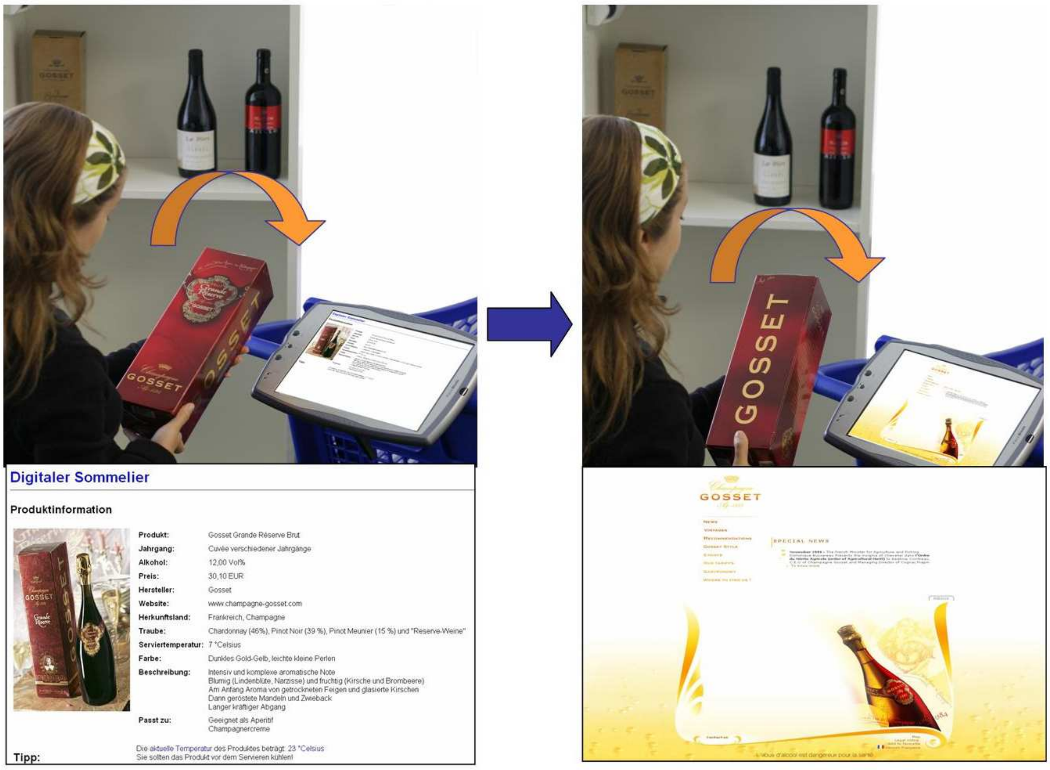


***Figure 8.7**: Interacting with a bottle of champagne*

8.2.2 DigiSom Sensor Processing

There are two independent tasks for the integrated sensors in the DigiSom application: Collecting and storing sensor readings that define the state of wines (temperature and vibration) using the μPart nodes, and acceleration sensors on the Crossbow motes to detect, whether a user looks at the front or at the rear side of a product box. Figure 8.8 shows the setup of the sensor processing filters used for the DigiSom.

Temperature processing was a rather straightforward task, there were only two minor issues which have to be dealt with. First, the μParts do not allow to set the sampling rate of temperature and vibration sensor independent of each other. To provide adequate reaction times in detecting vibration, we had to choose a sampling interval of about one second. This results in a steady stream of equal temperature values, which can be limited by the *NoRepeat* filter module. The second problem is that at the transition between two degree values, there is some interval of jittering, where sensor readings continuously switch between these values. To prevent these messages from slipping through the NoRepeat filter, we compute the mean of two successive readings. This is the simplest form of a moving average represented by the module *Mean* with a buffer size of two and with an overlap between buffers of one message (see Figure 8.8).

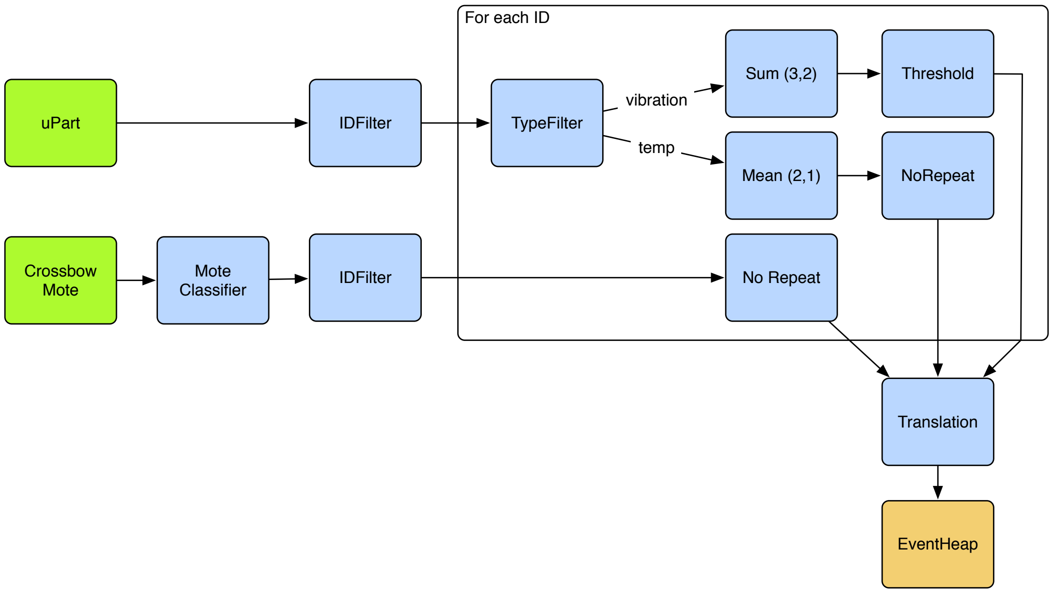


***Figure 8.8**: The filter chain of the DigiSom*

For movement detection the μPart nodes use an integrated ball switch, which enables the implementation of e.g. a wakeup on move function. But for deciding, if a bottle has been shaken or, for instance, just been moved, we have to further process the data stream of this sensor: Vibration readings are summed up over a buffer of three messages with an overlap of two. The result is then compared to a threshold, which represents the sensitivity of the detection process. To detect the orientation of the product boxes, we integrated Crossbow motes. Their two-axis accelerometer measures the acceleration caused by gravitation. A user looking at a product never holds it perfectly upright, but always in a certain angle such that she must not hold the box at eye level for a proper view. Taking this observation into account, we have trained our WEKA classifier to distinguish front, back and upright orientation.

After these preprocessing steps messages are translated to map sensor node IDs to products and published as EventHeap tuples to make them available to other parts of the DigiSom. The overall architecture of the system builds on the prototyping structure introduced in Section 5.2 and is visualised in Figure 8.9.

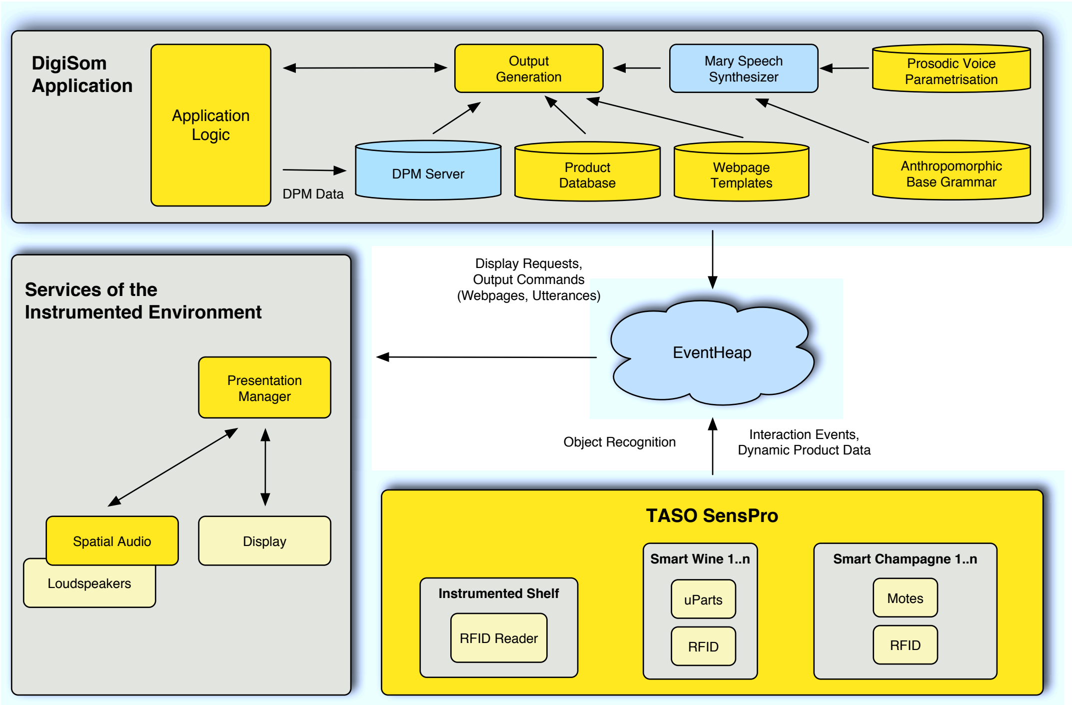


Figure 8.9: *The overall architecture of the DigiSom system based on the TASO Platform*

8.3 The Anthropomorphic Cocktail Shaker

The so-called Virtual Cocktail Guide is an interactive installation teaching users to mix virtual cocktails with real artefacts under the guidance of a virtual character. The overall installation consists of an instrumented bar counter, a large screen that displays the character, an instrumented bottle and an instrumented cocktail shaker. A set of RFID-tagged cards represents the collection of cocktails from which the user can chose the beverage that will be prepared by following the instructions of the virtual bar mixer. The system can be classified as an edutainment installation that guides users through the preparation of cocktails in a game-like fashion, while at the same time various aspects of cocktail mixing are conveyed during the course of the interaction. The installation is designed for a long term exhibit in a supermarket and also offers opportunities for introducing new products (i.e. *cross-selling*) to customers in the food and beverages domain. The overall setup is similar to the A.I. poker (as in Section 3.2.3.2) and the Virtual Constructor (see Section 3.2.3.1), but the VCG also incorporates tangible interaction with an animalistic smart object, the Anthropomorphic Cocktail Shaker.

The shaker implements animalistic behaviour by employing different types of affective sounds in reaction to certain user behaviour. The standalone version also adds speech output in order to comment on situations, which would be the task of the virtual cocktail guide in the original scenario. The implementation of the interactive shaker as a standalone life-like object has been realised as part of this dissertation and will be described in the following.

8.3.1 Interacting with the Anthropomorphic Cocktail Shaker

The interaction with the shaker takes place when the user is about to start with the actual shaking process, thus, after all ingredients have been filled into the shaker. When the shaker is assembled and thus closed, the user can start shaking the ingredients in a particular manner to finish the virtual cocktail.

Figure 8.10 shows the top level of the interaction flow of the standalone shaker modelled with the SceneMaker tool. The application starts with a brief introduction, explaining that the user is about to mix a drink and that all ingredients have already been virtually filled into the shaker. The shaker advises the user to close it, such that the user can start mixing what is inside. If the user picks up the shaker before it is closed and plays with it in a way such that real liquids would be spilled, the shaker is *disgusted* and explains that the user is lucky to play with virtual liquids. If the user does not do anything at all, the shaker turns *bored* and reminds the user that it is still waiting. After the shaker has been closed, it is ready to be shaken and vibrates regularly if nothing happens. The shaker recognises when it's being hit (including knocking on its surface) and plainly reacts *startled*. Shaking activities are evaluated as right or wrong, while appropriate shaking is achieved by a particular tempo of motion along the vertical axis of the shaker, which will be acknowledged by bursts of *joy*. Wrong shaking makes the shaker *angry* such that it will explain how to shake properly. After some time of correct shaking, the shaker is *pleased* and states that everything is well prepared now.

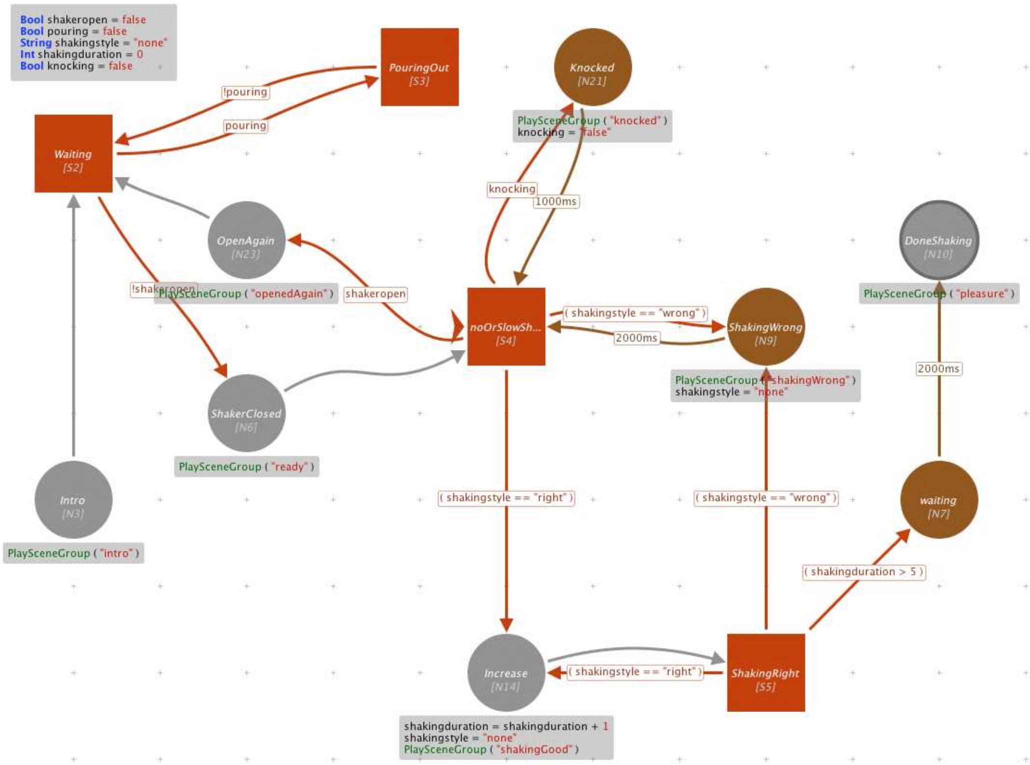


***Figure 8.10**: The top level of the scene graph for the anthropomorphic shaker*

Affective feedback is mainly represented by non-verbal sounds as described in Section 6.1.2, enriched with physiological sounds. The accompanying speech output is pre-recorded for the standalone version. All sounds can either be played back over the speaker integrated into the shaker or at the desktop computer hosting the overall application, which will result in higher audio quality when connected to a common hi-fi system (see Figure 8.11 for the final prototype of the Anthropomorphic Cocktail Shaker waiting for the user to start shaking).

In the following we will list all life-like elements integrated into the shaker interaction:

Physiological Feedback: During the introduction, all ingredients that have been virtually filled will be listed verbally, each of which followed by a *swallowing* sounds. When the shaker is closed and awaits the shaking, it imitates a *pulse* through vibration. In the end, when the correct shaking process is confirmed, the shaker will emit a *belch*.

Affective Auditory Icons: In order to reflect certain anthropomorphic states, the recorded and synthesised sounds introduced in Section 6.1.2 are rendered, representing *disgust*, *boredom*, *anger*, *joy*, being *startled* and *pleased*.

Expressing Personality: Following the guidelines introduced in Section 6.3, we created an anthropomorphic base grammar for an *extrovert* personality through pre-recorded phrases. The interaction flow also aims at not only being reactive to all user actions, but also at employing pro-activeness when the user appears to hesitate, which is a dialogue behaviour that further underlines a high degree of extraversion.

Figure 8.11: *The Anthropomorphic Cocktail Shaker*

8.3.2 Smart Object Instrumentation and Architecture

The cocktail shaker contains in its lowest compartment an invisibly embedded microcontroller platform (Arduino FIO[1]), powered by polymer lithium ion battery with 1000mAh, and connected to a series of sensors and actuators. It communicates wirelessly through a ZigBee module with a ZigBee counterpart connected to the PC that hosts the remaining software. The ArduinoFIO is connected to the following sensors and actuators:

ADXL335 +/-3g, breakout board: A 3-axis acceleration sensor to detect movements of the shaker.

Electret Microphone: A microphone to detect and measure acoustic events.

Light/Colour Sensor DJD-S371-Q999: A light sensor that can also measure RGB values of incoming light.

Sound Module SOMO-14D: A module that has a slot for a Mini-SD card, from which it can decode audio and render it on an integrated loudspeaker.

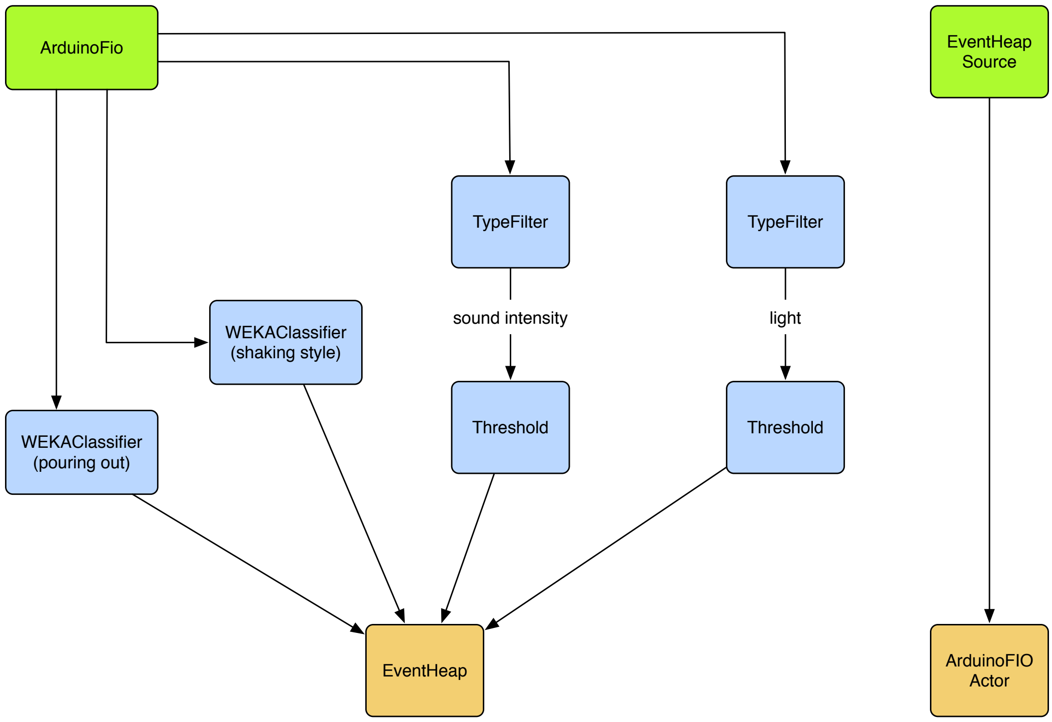


***Figure 8.12**: The sensor processing filter chain of the shaker*

Input recognition and controlling integrated actuation components is implemented with TASO SensPro (see Section 5.1): For recognition purposes, the shaker integrates acceleration sensors, a light/colour intensity sensor and a microphone. The light sensor is used to detect whether the shaker is open or closed, which is realised by checking the sum of the three RGB values against a threshold that can be adjusted depending on the lighting conditions of the environment. Due to its sensitivity it can detect light changes within the main compartment even through the separating layer of ABS thermoplastic. The microphone sensor data is checked to determine whether the shaker is hit by some thing or the user. The acceleration sensor data is used to infer whether the shaker is being held in a way such that liquids would flow out. It further provides data that is used to classify the shaking style, which could be right or wrong. Both actions are recognised through WEKA classifiers, applied on prerecorded training data. The sound module can trigger the direct playback of sound tracks stored on the Mini-SD card.

The basic setup of object instrumentation, sensor processing, interaction flow and vocal behaviour components is based on the TASO Platform developed in Section 5.2. TASO SensPro connects to the sensor board of the shaker, continuously receiving sensor data, which is processed and interpreted to generate events that are relevant for the interaction process. An extension to the SceneMaker listens to such sensor events and translates them into interac-

[1] `http://arduino.cc/en/Main/ArduinoBoardFio`, last visited June 30, 2010

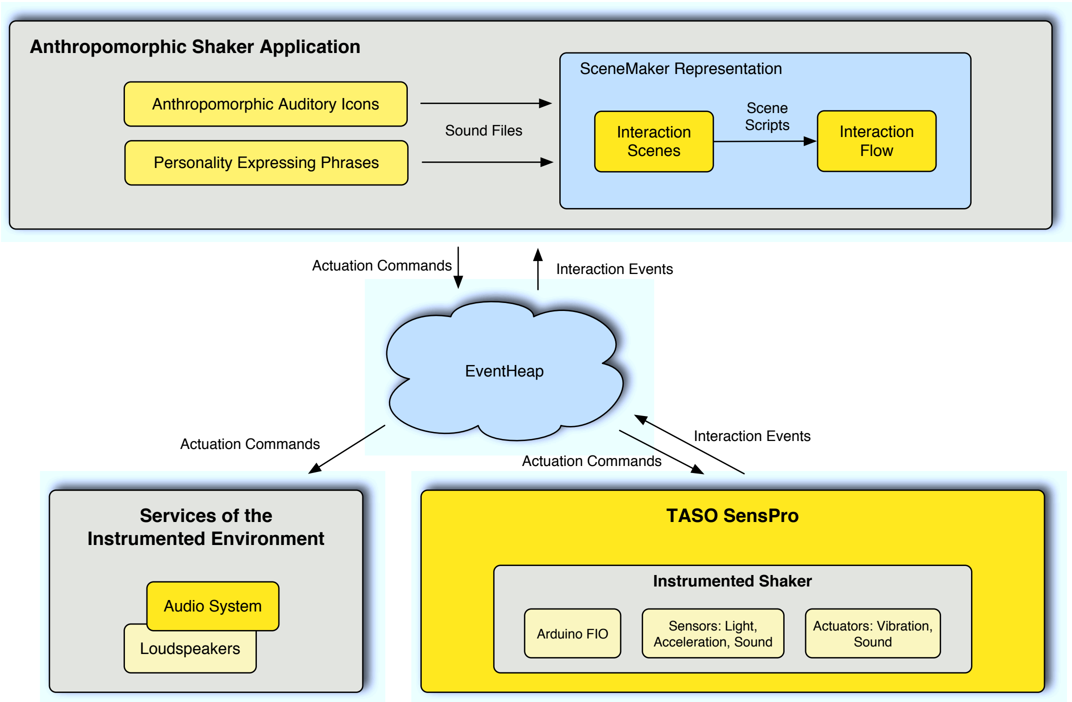


Figure 8.13: *The overall architecture of the Anthropomorphic Cocktail Shaker based on the TASO Platform*

tion events by calling corresponding functions of the SceneMaker representation, as outlined in Section 5.2. This module represents the main interaction logic and represents the states of the shaker and a description of scenes, which define the behaviour of the shaker. The SceneMaker representation generates *ActionSequences* that contain a series of timed, singular control statements that control audio output and vibration, which are fed back into the communication infrastructure and either rendered by a desktop computer or directly by the shaker. In this case, the actuation command initiated by our communication component of the SceneMaker instance are sent to the EventHeap, received by TASO SensPro and forwarded over the ZigBee connection to the ArduinoFIO board, which finally communicates with the sound module through binary signals. The sensor processing streams of the instrumented shaker are displayed in Figure 8.12.

Figure 8.13 visualises the structure of the architecture and the flow of data according to our prototyping environment.

8.3.3 Product Design Issues of the Instrumented Shaker

For the design and construction of the body of the shaker, we cooperated with a product designer (Ying Wang) who started with a series of sketches and 3D models to get first ideas and impressions of possible shapes and to have a ground for discussion. After a preselection and exclusion of certain design directions, the next iteration of the design process involved foam prototypes in original scales, which provided a more detailed impression of the haptics and spatial relations.

In addition to subjective aesthetic preferences, additional aspects such as ergonomics or technological constraints as discussed in Sections 4.1.5 and 4.2 have to be regarded in this process. In the following we summarise the partially conflicting aspects of the physical design process:

Shape and appearance: Aspects that relate to the general visual appearance of the object.

- The device should be instantly recognised as a cocktail shaker and hence display typical characteristics of this type of item.
- The general impression should resemble a soft and organic structure, such that the projection of anthropomorphism can occur naturally.
- The item should communicate basic affordances that reflect how to hold and shake it and how to open each of the two parts on top that allow to pour the contents out and to fill it with ingredients respectively.
- The ergonomics of the item have to enable the user to operate the shaker (holding, opening, closing, shaking) with minimal efforts.

Sensor integration: Integrating the microcontroller and required sensors poses certain constraints to the shape of the body (see Figure 8.14).

- The overall weight has to be balanced such that the item can stand firm on even ground.
- The electronic components have to hidden from a normal user of the shaker.
- Connectors to recharge the internal battery must easily be reached.
- The colour sensor needs visual contact to the shaker opening through the interior.
- The acceleration sensors need to be aligned such that the shaking classification can be reproduced.
- The integrated speaker should not be completely blocked by the case of the item.

The final model (displayed in Figure 8.15) incorporates these aspects and has been constructed out of ABS thermoplastic with a 3D printer.

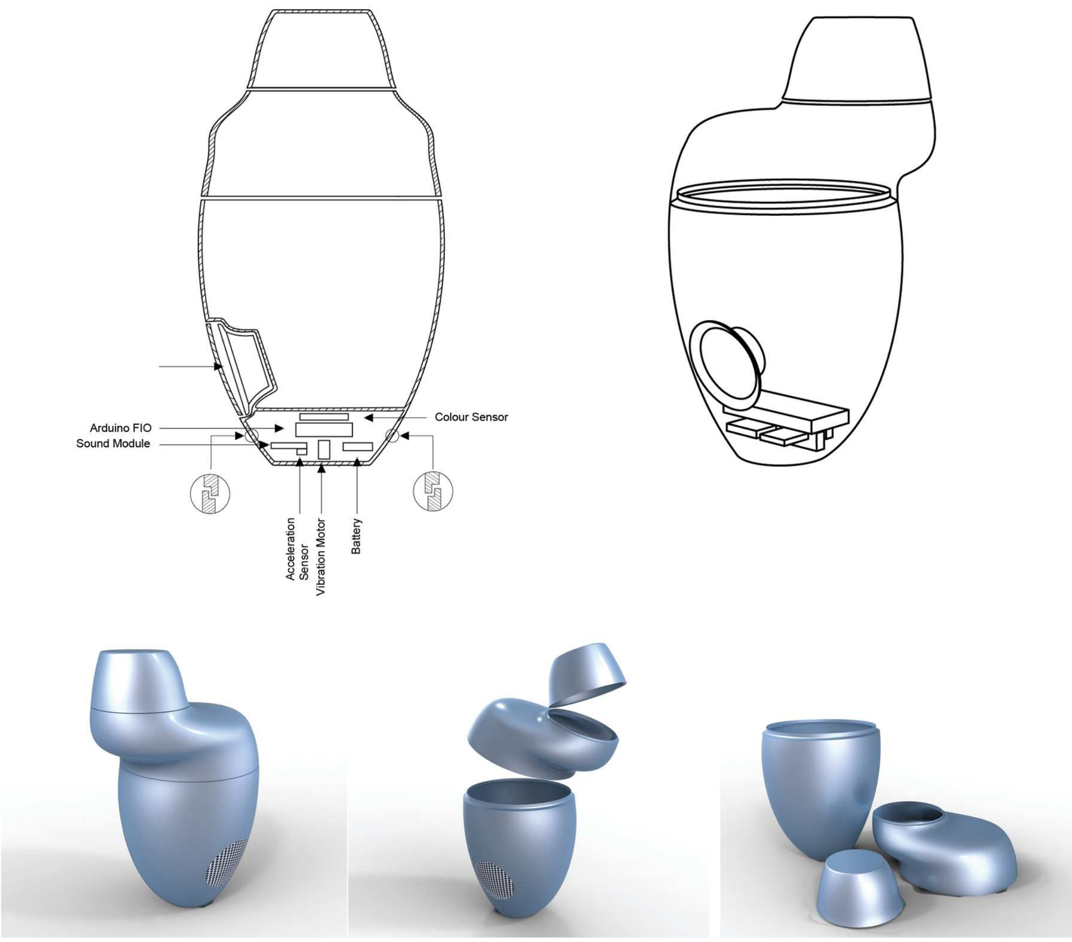


***Figure 8.14**: Technical drawings and 3D models of the shaker*

8.4 Synopsis

This chapter covered systems that exemplify the realisation of TASOs, demonstrating the possibilities of this type of smart object interface. The research area spanned by the TASO paradigm is very diverse and crosses the boundaries of a large number of research disciplines and application domains, such that the presented instances have to be seen as first probes into the design and research space of tangible interaction with anthropomorphic smart objects. The realised applications and concepts introduced in this chapter constitute a significant body of work outlining potentials and suggesting directions of future applications.

The Sensing Pill Box facilitates animalistic feedback to user actions that relate to the intake and storage of medical containers at home. It continuously checks environmental conditions and logs when it is opened and closed, inferring that the patient has retrieved a pill. Positive affective feedback affirms appropriate activities, while negative feedback warns

Figure 8.15: *The final prototype of the anthropomorphic shaker*

the patient. It's digital memory can further be inspected by other systems, e.g. for visualising its history.

The Digital Sommelier introduced in Section 8.2 shows how to apply talking objects in a shopping scenario by realising an interactive shopping assistant, which uses the concept of personalised speech to present multi-modal product information. Based on the prototyping framework introduced in Chapter 5, the application combines RFID- and embedded sensor technology to incorporate real time data about the product's state and user interactions with the product itself, in order to adapt the product presentations that will be given to the user. Furthermore, the data collected might also be used to feed and update digital product memories, that can be accessed by producers, retailers and end users. We also integrated our concept of talking products with personality (see Section 4.3.2) into this scenario: Each type of wine is associated with a distinct personality to represent a sense of brand and wine type. Personalities are reflected by prosodic attributes of the voice (see Section 6.2) and by different phrasing styles (see Sections 6.3 and A.2).

The Anthropomorphic Cocktail Shaker realises tangible interaction for virtual cocktail mixing, communicating with the user through synthesised affect bursts and recordings of physiological sounds that reflect states of the system and the interaction process. Verbal output, that underlines the extrovert personality, is generated as well, in order to enable the interactive shaker to provide more detailed explanations for the next step. We also presented the overall architecture of the system that makes use of interaction recognition based on various types of sensors, which are processed by TASO SensPro, the sensor-based interaction recognition framework introduced in Section 5.1.

9 CONCLUSIONS

This chapter concludes the dissertation with a summary of the scientific and practical contributions achieved during the course of this work. We will also outline opportunities for further research, based on the findings presented in this thesis.

9.1 Scientific Contributions

This dissertation has introduced the novel concept of tangible interaction with anthropomorphic smart objects (TASO), which has been derived by combining tangible with anthropomorphic user interfaces in the domain of smart object interaction. This is the first approach that consequently conceptualised anthropomorphic interaction with physical objects. We have motivated the idea behind this smart object interface paradigm in the introduction and underpinned the rationale behind this concept throughout the review of related research and projects. We have formulated seven research questions in Section 1.1, each of which we have regarded in this dissertation, as shown in the following.

The concept of TASO is an interdisciplinary approach that stems from various disciplines and research areas, which we have reviewed and analysed in Chapter 2, identifying relevant key aspects and leveraging them to a new domain and context. Defining and delimiting the opportunities and restrictions of smart objects and instrumented environments for post-WIMP interfaces, we conceptualised the domain of our work and identified anchors for this interaction concept. We have pointed out the role of prevalent interaction design aspects, aligned our approach to established TUI concepts, and discussed different layers of anthropomorphism in interfaces to provide a fundament for the TASO concept described in Chapter 4.

In order to benefit from current research efforts in relevant domains, we have reviewed a substantial body of related projects in Chapter 3 and identified their key attributes in the context of our TASO approach. This includes an analysis on a technological level as well, which informed the architecture of our prototyping framework introduced in Chapter 5. We also examined general findings of anthropomorphic and tangible interfaces and the results of first trials of other research groups, which explore smart object interfaces.

We have further derived design guidelines for TASO and presented technologies for effective prototyping of TASO instances. We also developed first use cases and novel approaches to realise selected TASO aspects, which have been evaluated in empirical user studies and constitute fundamental building blocks for future TASO realisations. The scientific results of this work are summarised as follows:

- **A novel interaction paradigm towards life-like smart objects** (concerns question 2 and 3)

  We have combined concepts of tangible interaction and anthropomorphism to construct a novel approach to interact with smart objects, which are expected to become integral parts of our everyday life. The tangibility of such physical objects is merged with the intuitive understanding of anthropomorphism in order to support particularly novice users or casual interaction scenarios. Both paradigms, tangible and anthropomorphic user interfaces have in common that they exploit established human skills and knowledge learned in everyday life. We have elaborated this concept in Chapter 4.

- **Guidelines for the design and construction of TASOs** (concerns question 1, 2, 3, and 4)

  The design and construction of TASO cross domain boundaries and require knowledge in several disciplines that relate to user interface design and prototyping. This work provides an exploration of the design space and introduces guidelines to support interface designers who consider to apply the interface paradigm introduced in this work. We have also developed classes of TASO that involve different levels of anthropomorphism, which suit to different application scenarios.

- **Expressing personality through prosodic parametrisation** (concerns question 7)

  One building block for the construction of TASO interfaces is the expression of personality of talking objects. We have investigated how prosodic parameters of synthetic speech can influence the impression on the speaking entity's personality. On top of a literature survey we conducted an empirical user study that shows how prosodic parameters can be modelled to express certain personality dimensions. Modelling personality exclusively with prosodic features has the advantage that the spoken text does not have to be altered and that it remains under full control at design- and runtime of an interactive system.

- **Personality expression in dialogue phrases** (concerns question 7)

  Personality is certainly not only expressed in qualitative attributes of a voice, other properties of a speech dialogue are also essential, such as the employed vocabulary or the general discussion behaviour. The development of phrasing and speaking behaviour guidelines for natural language dialogues in Chapter 6 complements the prosodic parametrisation of speech and provides additional means for reflecting personality in natural language interfaces of smart objects. We have also shown how to build an anthropomorphic base grammar, from which utterances can be selected during runtime, depending on the personality that is intended to be reflected by the associated object.

- **Anthropomorphic auditory icons** (concerns question 7)

  Animalistic smart objects constitute a particular class of TASO that does not incorporate speech or personality but mimics behaviour that results from basic needs and desires corresponding to the concept of primitive psychology. Expressive means of this class involve sounds that we have called anthropomorphic auditory icons, which generally reflect states and reactions of life-like, animalistic entities. We have constructed a series of such sounds and introduced a method of generating new sets of such affective sounds for other objects. We also showed the effectiveness of the recognition of affective states in anthropomorphic auditory icons in an initial user study.

In addition to these conceptual results, several practical contributions have been made particularly to the domain of prototyping TASO in instrumented environments:

- **TASO Platform** (concerns question 5 and 6)

  During the realisation efforts of TASO instances we have built and refined a prototyping framework that allows rapid prototyping of smart object based interfaces that employ tangible input recognition and anthropomorphic system behaviour. The modular overall architecture respects the dynamic interdependencies between smart objects and instrumented environments and allows for flexible integration of hardware and services embedded into objects and shared by the environment, based on asynchronous communication via a shared tuple space.

- **TASO SensPro** (concerns question 6)

  We further developed a programming library for sensor-based prototyping of smart objects, which enables rapid development with reusable code fragments in stream-based processing structure. The library incorporates a variety of off-the-shelf sensor systems and allows for the realisation of smart object prototypes with minimal amount of code, including the recognition of complex user input such as gestures based on acceleration data through the integration of machine learning algorithms provided by the widespread WEKA toolkit. In addition to that, a graphical user interface enables even non-experts to prototype smart object applications with a node-based interface.

- **Supportive environmental services** (concerns question 5)

  In many cases smart objects are not isolated but interact with or even depend on instrumented environments, which also host technologies and services, from which smart object systems can benefit. We have implemented a series of such services to support the prototyping of TASO scenarios, comprising software for the realisation of spatial audio (SAFIR), a visual user tracking component (MaMUT) and a central service that manages and controls shared output devices in distributed environments (PM).

9.2 Scientific and Public Outreach

A number of aspects covered by this work have been published in leading international conferences and workshops. We summarise the contributions in the following in chronological order:

- International Conference on Intelligent Environments, Athens, 2006
- Mensch und Computer, Gelsenkirchen, 2006
- Speech Prosody, Dresden, 2006
- Workshop on the Design of Smart Products, Furtwangen, 2007
- International Conference on Intelligent User Interfaces, Honolulu, 2007
- International Conference on Intelligent Environments, Ulm, 2007
- International Conference on the ”Internet of Things”, Zurich, 2008
- International Conference on Tangible and Embedded Interaction, Cambridge, 2009
- IADIS Interfaces and Human Computer Interaction, Freiburg, 2010
- International Conference on Tangible, Embedded, and Embodied Interaction, Funchal, 2011

Particularly the Digital Sommelier has been exposed to the public on various exhibitions, since it was the first instance that has been realised in the context of this work. This includes the *CeBIT* Fair in Hannover (2008, 2009), *”Deutschland und China - Gemeinsam in Bewegung”* in Guangzhou (2008), Shenyang (2009) and Wuhan (2009), China, *”Bürgerfest der Länder”* in Berlin (2009) and *”Land der Ideen”*, Zentrale Globus SB-Warenhaus Holding in St. Wendel (2009). In addition to that broadcasts on radio, TV and other media have reported on this prototype, indicating a high public interest in the ideas behind this project.

9.3 Opportunities for Further Research

The main focus of this dissertation was to explore concepts and technologies that allow for the realisation of an anthropomorphic interaction paradigm for smart objects utilising tangible input. The resulting interaction concepts, design guidelines, prototyping tools, and building blocks for anthropomorphic expression provide the means to realise a large variety of extensions and variations in an abundance of facets. The implemented prototypes that build on these results demonstrate the potentials but can certainly only be seen as initial probes into a highly complex and diverse design and research space. There are several areas that can build upon the results of this thesis and pick up aspects for further research:

Multi-object interaction: A topic to consider in future research efforts is that of interacting with a multitude of anthropomorphic objects. So far the multi-object scenarios focused to one-to-one interaction, where objects do not consequently interact or connect to each other on a functional or interactive level. But the natural and object-centric interaction paradigm presented in this thesis provides promising means to regard such complex interaction scenarios, which could be elaborated in appropriate use-cases, e.g. in shopping contexts that involve products that compete against each other when the consumer requires support in shopping decisions. This also provides opportunities for influencing and guiding the user towards certain products or to follow up with cross-selling recommendations.

Investigating the possibilities of new materials: Smart materials and miniaturised shape changing actuators for instance based on shape memory alloys are just before market breakthrough and provide additional means for new generations of objects. It allows to integrate additional affordances for TASOs and organic user interfaces in general. The advent of OLED technologies promises opportunities to enhance certain instances of smart objects, e.g boxed products, in order to utilise the visual channel as well, for instance by rendering faces and other body parts, which could extend the TASO concept by providing additional means for expressing life-likeliness in physical objects.

In-depth user studies: We have developed and evaluated several corner stones for further examinations of new TASO instances, which again have to be evaluated with real users in realistic scenarios. Such extensive user studies will reveal substantial insights about user acceptance of and attitudes towards TASOs, which will help to refine and enhance the proposed concepts.

Continuation of fundamental research in related disciplines: The presented results also have an impact on ground work in research areas beyond the core themes involved in the work at hand and provide new anchors for continuing research streams started by this dissertation. For instance in the realm of auditory interfaces, where synthesised affective sounds are yet almost unexplored. There are several possibilities to proceed in order to create a deeper understanding of the attributes of sound and how they relate to emotions. A logical next step could also be to parametrise anthropomorphic auditory icons for more flexibility and customisation possibilities in auditory interfaces or to apply naturalistic carrier sounds, such as a dog bark or a birdsong. Our efforts in exploring the connection between prosodic parameters of synthesised speech and personality perception also constitute first steps into new territory and can be continued by e.g. exploring how to reflect more complex personality profiles through prosody in speech.

These research directions will not only affect their respective fields but inform various disciplines that relate to intelligent user interface, such as human-robot-interaction.

Exploration of further use cases: The medical scenario introduced in Section 8.1 will be extended, focusing at the improvement of the patient's compliance over longer terms of use. For this thesis we concentrated on short term and casual interaction in selected scenarios, but as outlined in Chapter 4, we also introduced the class of bonding objects in an anthropomorphic object scenario. As suggested, such bonding smart objects can potentially facilitate persuasive interfaces, particularly over a longer term of interaction, which requires longterm user studies accordingly. This concept is particularly interesting for such a medical use case, since the overall motivation is to influence and improve the patient's behaviour according to the doctor's prescription. We believe in the strong potential of bonding objects in this and similar use cases that benefit from increased motivation of users and which involve frequent and long term use. Our approach could hence theoretically improve on this complex of problems around patient's compliance even on a subtle interface level.

Similarly, the HCI domain of human-car-interfaces provide another ground for exploration of recurring and long term interaction. In-vehicle assistance systems are often used on a daily basis, providing the required conditions to control and amplify a social relation between car and user and exploit the proposed concepts for persuading and influencing the driver to e.g. adhere to speed limits or to induce other behaviours that relate to security or energy consumption.

The DigiSom already indicates the potentials of TASOs in retail environments. Further prototypes might aim at influencing consumer decisions by deliberately assigning particular personalities to certain objects, depending on consumer profiles or learning from the analysis of consumer behaviour that relates to reactions on certain personalities.

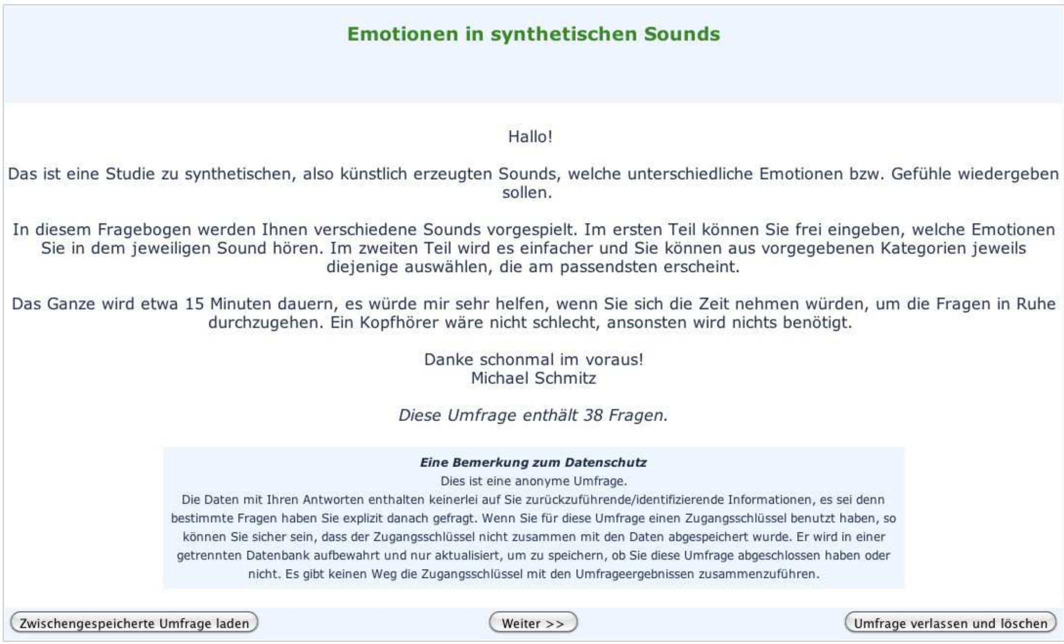
The concepts and prototyping tools presented in this dissertation facilitate the realisation of additional prototypes and support the involved researchers, developers and designers on multiple levels.

A APPENDIX

A.1 Anthropomorphic Auditory Icons - Evaluation Data

A.1.1 Questionnaire

The user study was conducted in German carried out online by German native speakers. The following pictures show screenshots of the questionnaire webpages.


Emotionen in synthetischen Sounds

Hallo!

Das ist eine Studie zu synthetischen, also künstlich erzeugten Sounds, welche unterschiedliche Emotionen bzw. Gefühle wiedergeben sollen.

In diesem Fragebogen werden Ihnen verschiedene Sounds vorgespielt. Im ersten Teil können Sie frei eingeben, welche Emotionen Sie in dem jeweiligen Sound hören. Im zweiten Teil wird es einfacher und Sie können aus vorgegebenen Kategorien jeweils diejenige auswählen, die am passendsten erscheint.

Das Ganze wird etwa 15 Minuten dauern, es würde mir sehr helfen, wenn Sie sich die Zeit nehmen würden, um die Fragen in Ruhe durchzugehen. Ein Kopfhörer wäre nicht schlecht, ansonsten wird nichts benötigt.

Danke schonmal im voraus!
Michael Schmitz

Diese Umfrage enthält 38 Fragen.

Eine Bemerkung zum Datenschutz
Dies ist eine anonyme Umfrage.
Die Daten mit Ihren Antworten enthalten keinerlei auf Sie zurückzuführende/identifizierende Informationen, es sei denn bestimmte Fragen haben Sie explizit danach gefragt. Wenn Sie für diese Umfrage einen Zugangsschlüssel benutzt haben, so können Sie sicher sein, dass der Zugangsschlüssel nicht zusammen mit den Daten abgespeichert wurde. Er wird in einer getrennten Datenbank aufbewahrt und nur aktualisiert, um zu speichern, ob Sie diese Umfrage abgeschlossen haben oder nicht. Es gibt keinen Weg die Zugangsschlüssel mit den Umfrageergebnissen zusammenzuführen.

Zwischengespeicherte Umfrage laden | Weiter >> | Umfrage verlassen und löschen

***Figure A.1**: Introduction page*

Figure A.2: *Announcing personal questions*

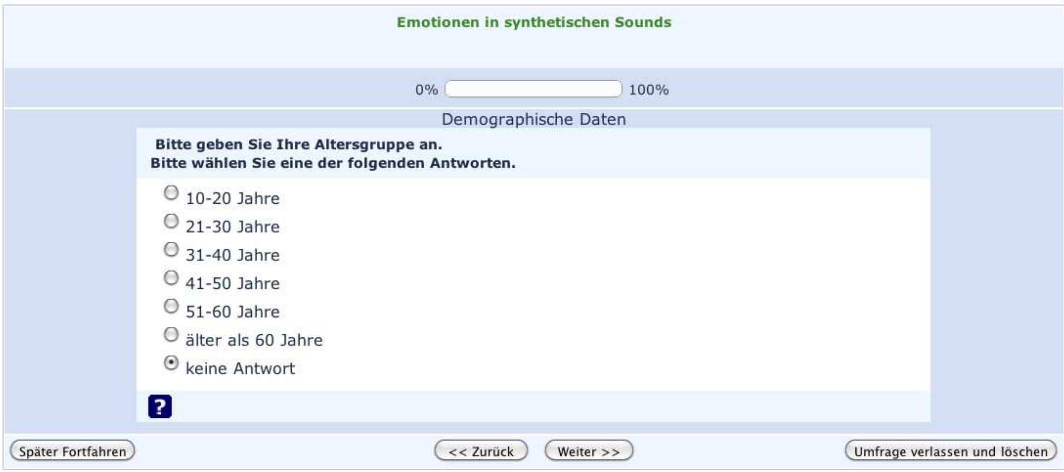


Figure A.3: *Inquiring the age range*

Figure A.4: *Inquiring the gender*

***Figure A.5**: Test page to let participants check the technical requirements*

***Figure A.6**: One page of the first part of the study with free text entry*

***Figure A.7**: Announcing the second part of the questionnaire*

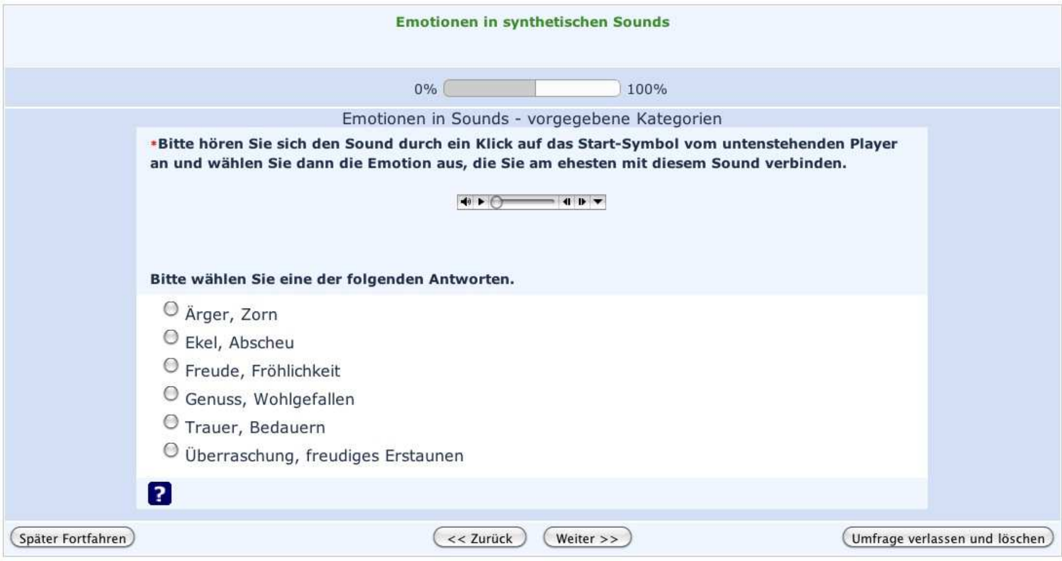


Figure A.8: *One page of the second part of the study with given emotion categories*

A.1.2 Qualitative Data

Table A.1 lists all participants, their gender, age range and whether they have completed the study.

User ID	Gender	Age range	Completed study
4	male	21-30	Yes
6	female	21-30	Yes
7	male	31-40	Yes
8	male	21-30	Yes
11	male	21-30	No
13	female	21-30	Yes
14	male	31-40	Yes
16	male	31-40	Yes
17	male	21-30	Yes
19	male	21-30	No
22	male	31-40	No
23	Not specified	41-50	No
24	male	21-30	No
25	male	31-40	No
28	male	21-30	Yes
30	female	31-40	Yes
31	female	31-40	Yes
36	male	21-30	Yes
38	male	21-30	No
41	male	41-50	No
42	female	31-40	No
45	male	21-30	No
46	female	31-40	Yes
47	male	31-40	Yes
48	male	31-40	Yes
49	male	not specified	Yes
51	not specified	not specified	Yes
53	male	31-40	Yes
54	male	31-40	Yes

Table A.1: *Overview on participants and demographic data*

In the following we list all comments made for each sound.

ID	Sawtooth_Happy	Sawtooth_Pleasure	Sawtooth_Surprise
4	freude	mhhh, lecker	überraschung
6	nervige tussi an der theke neben einem	klugscheisserisch	wieder klugscheisserisch
7	Linkische Freude	gutmütige Zustimmung	Wiederwille
8	Lachen	genervt	wütend/schimpfend
11	Schadenfreude		Erkenntnis
13	Auch Schadenfreude	So ein bisschen wütend, auch wieder ermahnend	Anerkennung, also z.B. wenn man was erzählt bekommt und dann merkt, dass die andere Person recht hat...
14	fröhlich , lachend	bewundern	interessiert, zustimmend
16	hämisches lachen	langeweile	trotz
17	Freude, Häme	Mürrisch	Erstaunen
19	schadenfroh froh	befriedigt wohlbefindent entspannt	interessiert/verstanden
22	fröhlich	stark; neutral; was sind den die auswahlsmöglichkeiten überhaupt?	überraschung.
23			
24			
25			nervig
28	freude		verstehen interesse
30	spontanes Amüsiert-Sein	Missfallen, von jemandem oder etwas gestört oder unsanft geweckt werden	(freudiges) Erstaunen
31	gehässig lachend verhöhnend	bedrohlich dominierend	genervt cholerisch
36	Lachen	Erstaunen	Erstaunen, Erleuchtung, Idee
38			
41	grinsen		
42			genervt
45	:::	...	pi pa po
46	Lachen, Freude	Zufriedenheit	freudige Überraschung, guter Einfall, Staunen, Freude
47	freude	ekel	erstaunen
48	Schadenfreude	Ärger	Ärger
49			
51			
53	idiotie freude verzweiflung	einsicht verstehen erleichterung	verärgert pikiert trotzig
54	Schadenfreude	lecker, sehr bestimmend	positiv überrascht

***Table A.2**: Comments on sawtooth-based positive emotions*

ID	Sawtooth_Anger	Sawtooth_Disgust	Sawtooth_Sad
4	wut, ärger	ekel	schmerz, frust
6	Bauernhof	ekelhaft, ablehnend, wie wenn jemandem was nicht schmeckt	morgens nicht aufstehen wollen
7	aufgeregt sein	missmutig sein	entspannt sein
8	Schreien	verärgert	nachdenklich
11		Generftheit, angewiedert sein	Müdigkeit
13	Jubel, Freude, also ein Jubelschrei	Ekel, Generve, Beschwerde	Trägheit, Faulheit, Widerwille
14	wütend	beleidigt	enttäuscht
16	schreck	aufmerksamkeit heischen	nichts
17	Schmerz, Irritierung	Enttäuschung	Enttäuschung, Schmerz, Verzweifelung
19	fallen ärgern	ekelhaft	abwägend
22		abstossend, yuck.	
23			
24			
25			
28	angst	ärger enttäuschung	
30	Wut, Schreck	Ekel, Wiederwillen, Trotz	Widerwillen
31	genervt anklagend	beleidigt rückzügig	jammernd aber mit witz in der tasche, sich selber lustig nehmen könnend
36	Beschimpfen, jemanden erschrecken	Ekel	Schmerz
38			
41			
42	Agressionen		
45			swaugouh
46	Ärger, Wut	Ablehnung, Ärger	Enttäuschung, Weinen
47	Angst	ekel	Niedergeschlagen
48	Wut	Ekel	Ärger
49			
51			
53	ärger schmerz	ekel	beenden zerstören
54	Zorn	ekel	enttäuscht

Table A.3: *Comments on sawtooth-based negative emotions*

ID	Sine_Happy	Sine_Pleasure	Sine_Surprise
4	ängstlich	mhhh, lecker	erstaunen
6	jippieh, party on, es geht raus auf die strasse	angenehm, vorfreudig	oh, das find ich aber gut!
7	flatterhaft sein	Erstaunt sein	ängstlich sein
8	weinen	Nachdenken	Staunen
11		begreifen	Verwunderung
13	Unstetigkeit, Ratlosigkeit, eingeordnet in Farben wäre das ein "blauer" Ton	wohlfühlend	Überrascht
14	ängstlich	genießend	verwirrt
16	erstaunen oder erschrecken	enttäuschung	verneinung
17	Ungewissheit, Mysterium	Enttäuschung und Schmerz	Spannung, teilweise Freude
19		fein schmeckend lecker	drehen
22			
23			
24	unsicherheit, spannung		
25			
28		verwunderung ratlosigkeit	
30	Schadenfreude	genervt sein	leichtes Erschrecken, Überraschung
31	zart singend fein fröhlich	traurig fragend	aufbrausend outgoing
36	Schüchternheit	Erstaunen	Erstaunen
38			
41		überlegen, lachen,	
42			
45	:9´)	mmm	
46	Freude, Lachen	Überraschung, Staunen	Behaglichkeit
47	ekel	Mist	erstaunen
48	Schadenfreude	Ärger	Ärger
49			
51			
53	erregung freude begeisterung	überaschung	erregung schwingung
54	fröhlich	lecker, schmeckt, gefällt	entspannt

***Table A.4**: Comments on sine-wave-based positive emotions*

ID	Sine_Anger	Sine_Disgust	Sine_Sadness
6	yipiehh	eieiei,...	langweile
7	nun komm schon!	nun komm schon (etwas vorwurfsvoller)	unentschlossen
8	freudiges Erkennengeben	Unsicherheit	traurige Hilflosigkeit
11	besorgtes Rufen	bibbern vor kälte	nachdenklich
13			Gleichgültigkeit
14	Kam dieser Sound nicht schonmal vor? Traurig, rumheulend	Erstaunen, Verwunderung, überrascht sein	nachdenklich
16	aufgebracht	fragend, nicht wissend	traurig
17	überraschung	bis hierher und nicht weiter	enttäuschung
19	Angst	Überraschung	Enttäuschung und Versagen
22			bad beat
23	fröhlich; wie ein Roboter vom Starwars	froehlich	traurig
24	verschlagen		müde, resigniert
25			
28			urwald-feeling
30			enttäuschung
31	Angst, Unterwürfigkeit	Ekel	Frieren
36	heulend klagend jammernd	zweifelnd beweglich	melancholisch versunken warm
38	Anstrengung	Entspannung	Entspannung
41		anstrengend	
42		komische Eule, kein Bezug	
45			
46	...		...
47	Angst	Angst, Überraschung	Ärger, Unzufriedenheit
48	trauer	Staunen	trauer
49	Angst	Langeweile	Langeweile
51			enttäuschung
53			
54	hilfegesuch ansprache aufmerksamkeit auf sich lenken	aufmerksamkeit bestätigung	entäuscht versagen falsch
	hilfesuchend aber gut gelaunt	fragend, resigniert	müde

Table A.5: *Comments on sine-wave-based negative emotions*

ID	Square_Surprise	Square_Happy	Square_Pleasure
4	erstaunen	freude, lustig	lazy
6	bedauern	schadenfroh	?
7	Überascht sein	Zuversicht	Zustimmen
8	Staunen, aber positiver (=WOW) Kompliment	mitlachen	ablehnende Haltung
11	unklare Erkenntnis		
13	Verärgert	Schadenfreude	Stress
14	sauer auf jemanden	meckern, tadeln	keine Idee?!
16	nichts	meckerndes lachen	empörung
17	Freude, Überraschung	Freude und Häme/Erniedrigung	Verspannung, Ungerreimtheit
19	zusprechend interresiert annehmend	lachen	
22	fröhlich; elefantastisch	komplex; neutral vielleicht.	
23			
24			
25			
28	vderstehne	belustigung	
30	genervt/gestört von oder sauer auf jemanden/etwas sein	Schadenfreude	Erkenntnis, Ausdruck dafür, etwas (endlich) verstanden zu haben
31	fröhlich wirbelig	meckernd genervt	unheimlich bedrängend
36	Verständnis, Idee	Schadenfreude	Erleuchtung
38		rastlos	
41			
42			
45	zog zog	...	
46	Frage, Unsicherheit	Unruhe	Ruhe
47	überraschung	Freude	ärger
48	Überraschung	Freude, Schadenfreude	Ärger
49			
51			
53	missachtung	schadenfreude freude	falsch fehler
54	entzückt	witzig, fröhlich, ein wenig schadenfroh	überrascht aber ein wenig enttäuscht, depressiv

***Table A.6**: Comments on squarewave-based positive emotions*

ID	Square_Anger	Square_Disgust	Square_Sadness
4	wut ? ärger ?	moment mal	erstaunen
6	knie aufgeschürft	lass mich in ruh	ohhh nicht schooooon wieder
7	alarmiert sein	fahrlässig sein	Hilflosigkeit
8	ängstlich	zorniges Rufen	Schmerzen
11	Unzufriedenheit		Gelassenheit
13	Das ist jemand alarmiert, vielleicht geschockt, auch ein bisschen überascht und verwundert	Ermahnend	Beeindruckt
14	alarmiert sein	aufgebracht	zustimmend, genießend
16	warnung	empörung	aua!
17	Aufgewühlt, unter Spannung, in Alarmbereitschaft	Überraschung	Erstaunt, Erregt
19		überrascht	zufrieden entspannt
22	aufgeregt, wie ein elefant die sein russel benutzt		
23			
24			
25			
28			
30	Missfallen, Angst	Beschwerde	Erleichterung
31	redselig ausschweifend	flüchtig, nicht bleibend ungeduldig	staunend abwertend
36	Angst, Erschrecken	Bestätigung	Erstaunen
38			überraschte Erkenntnis
41			
42			
45			---
46	Überraschung	Zufriedenheit, Freude	Unzufriedenheit, Trauer
47	Pech	überraschung	pech
48	Angst	Ekel	Ärger
49			ach schade
51			
53	schmerz aufschrei	gleichgültigkeit	wecken ende der kräfte
54	empört	überrascht	Schmerz

Table A.7: *Comments on squarewave-based negative emotions*

A.1.3 Quantitative Data

In this section we will present the results of the second part of the study, where participants could select one emotion category from a list to assign it to the presented sound. For each sound we will show a diagram with the summary of the participants' choices.

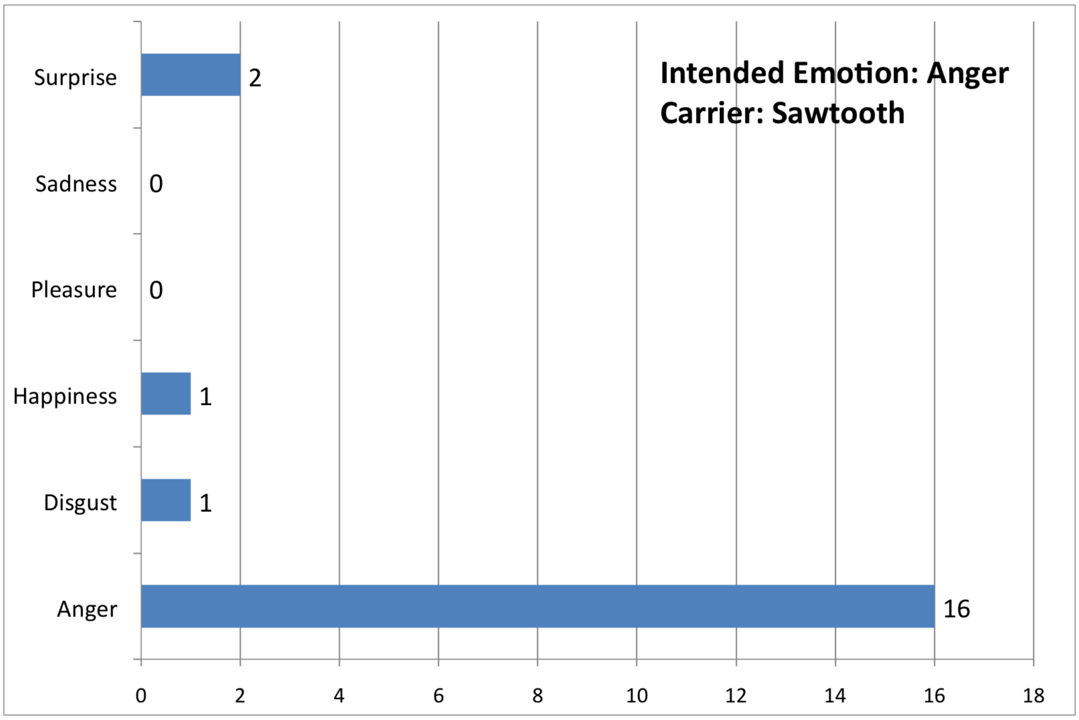


***Figure A.9**: Participant's ratings for anger based on sawtooth*

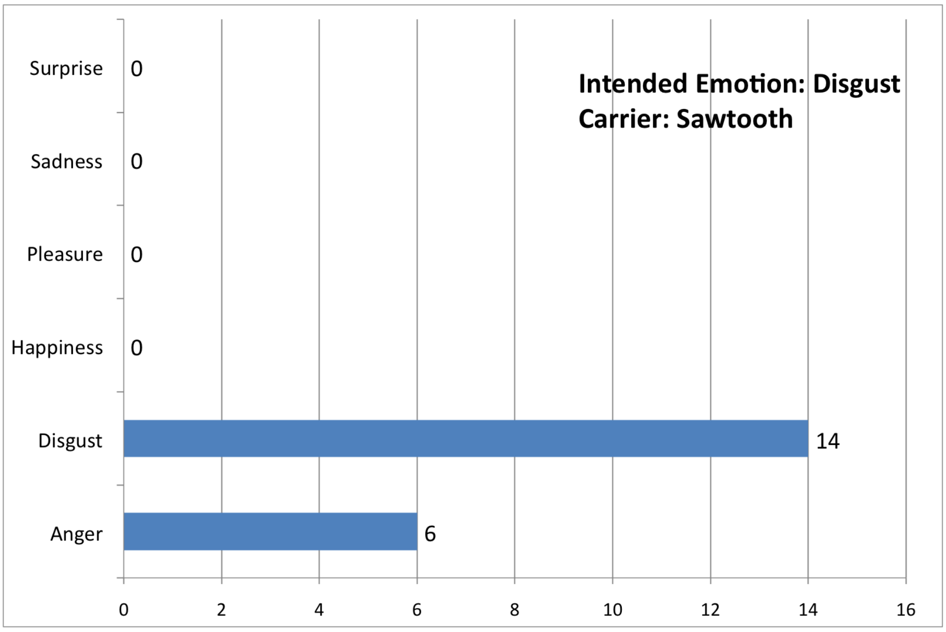


***Figure A.10**: Participant's ratings for disgust based on sawtooth*

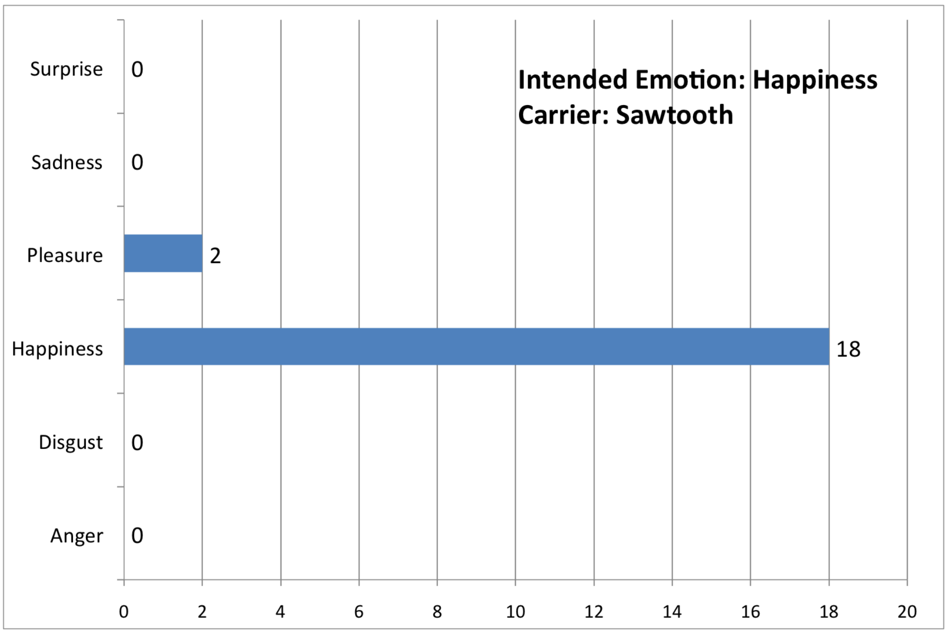


***Figure A.11**: Participant's ratings for happiness based on sawtooth*

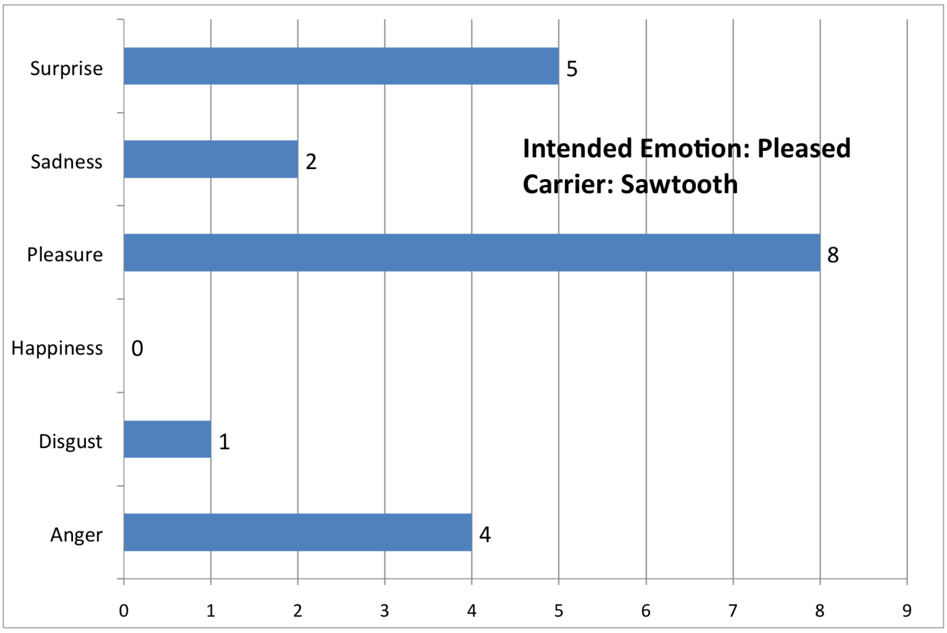


***Figure A.12**: Participant's ratings for pleasure based on sawtooth*

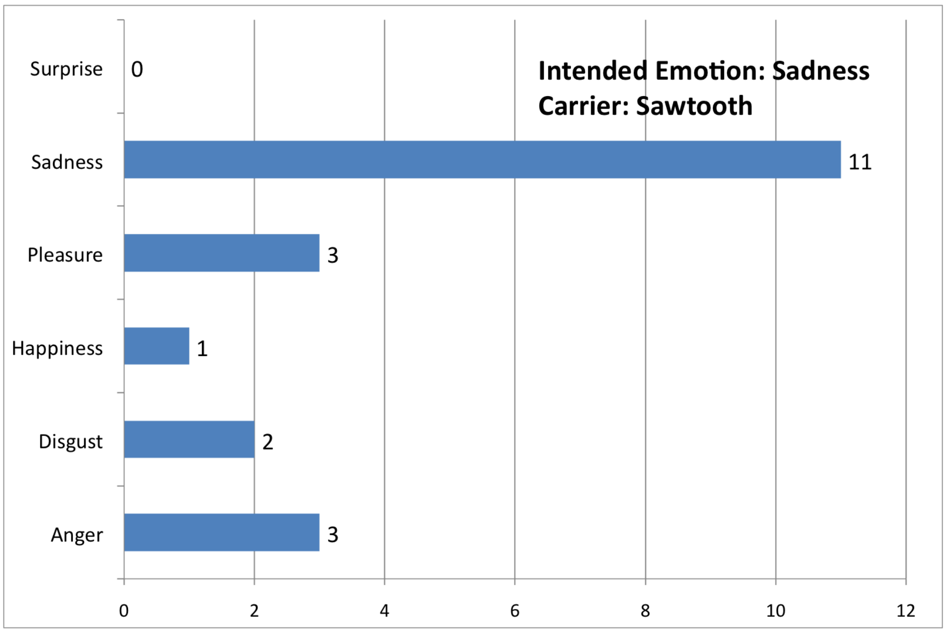


***Figure A.13**: Participant's ratings for sadness based on sawtooth*

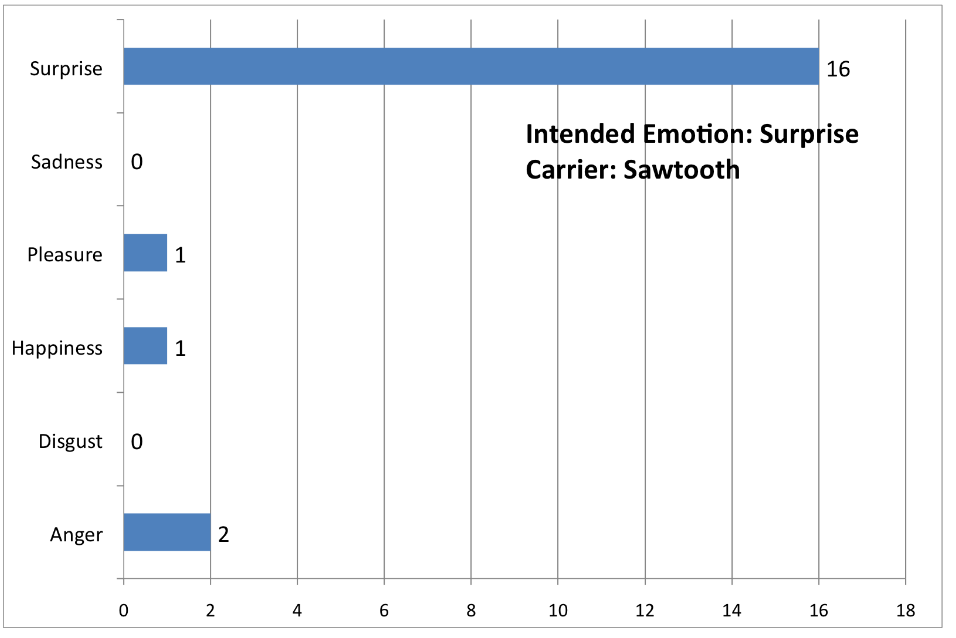


Figure A.14: *Participant's ratings for surprise based on sawtooth*

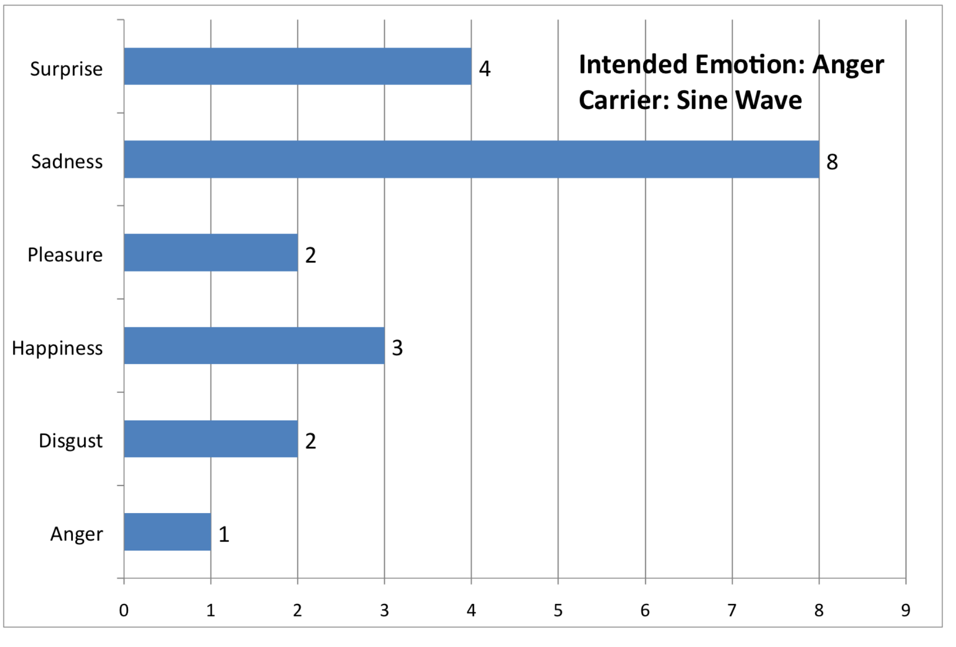


Figure A.15: *Participant's ratings for anger based on sine wave*

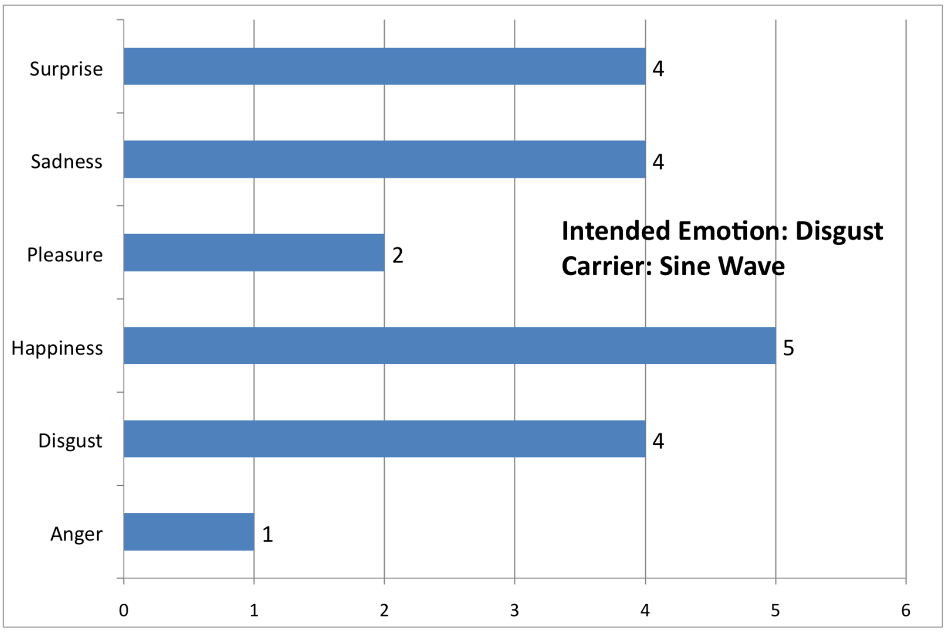


Figure A.16: *Participant's ratings for disgust based on sine wave*

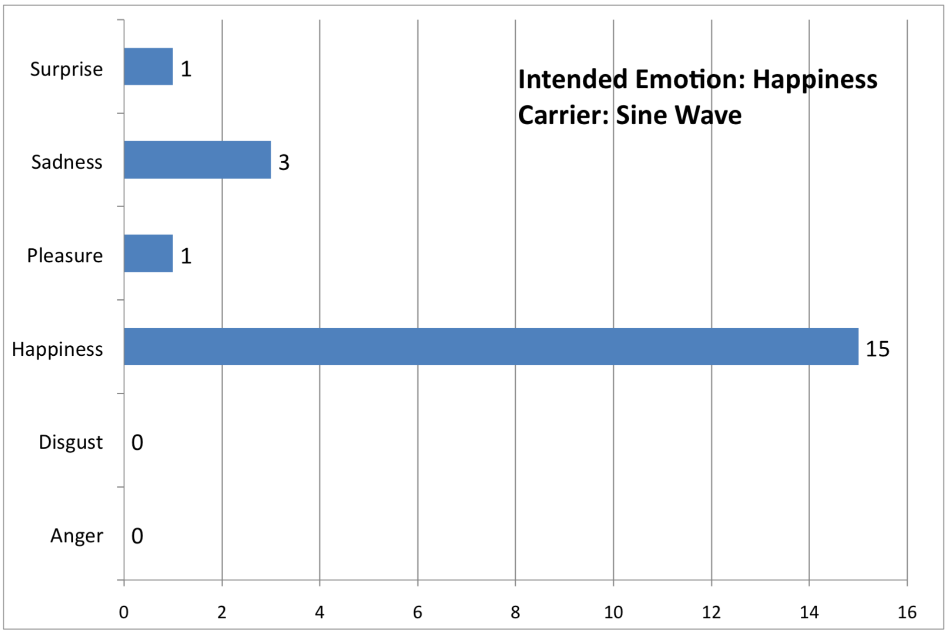


Figure A.17: *Participant's ratings for happiness based on sine wave*

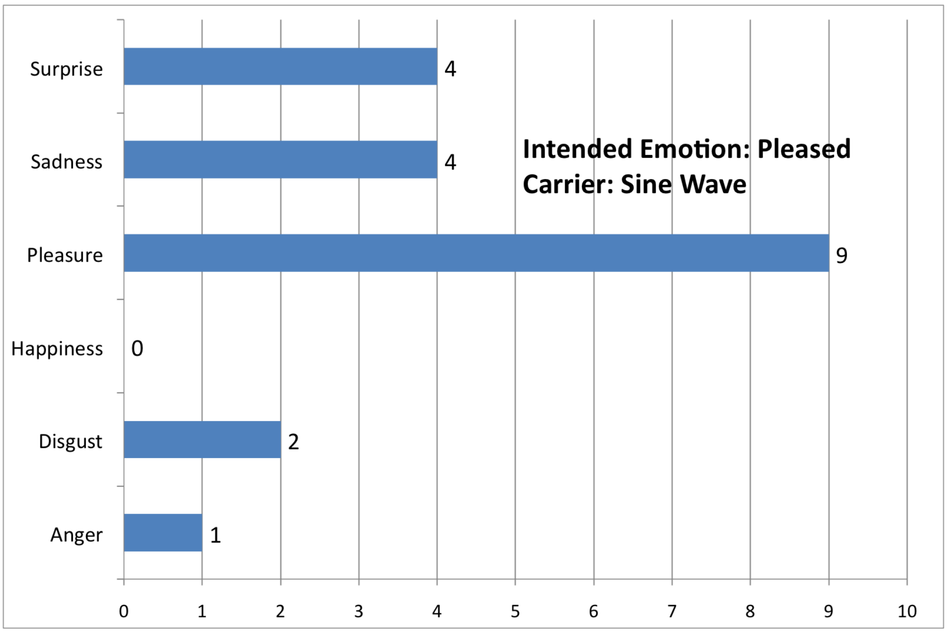


Figure A.18: *Participant's ratings for pleasure based on sine wave*

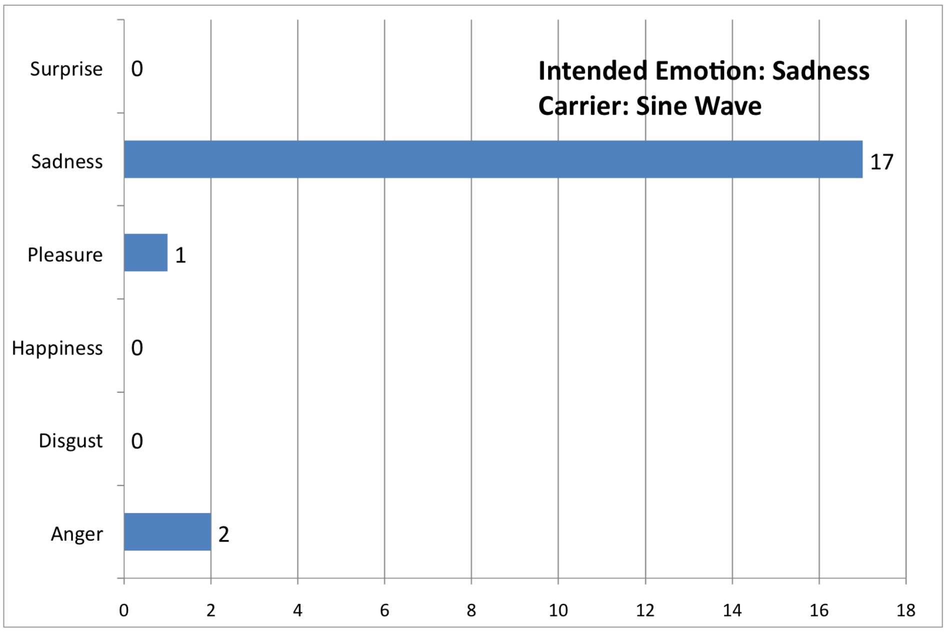


Figure A.19: *Participant's ratings for sadness based on sine wave*

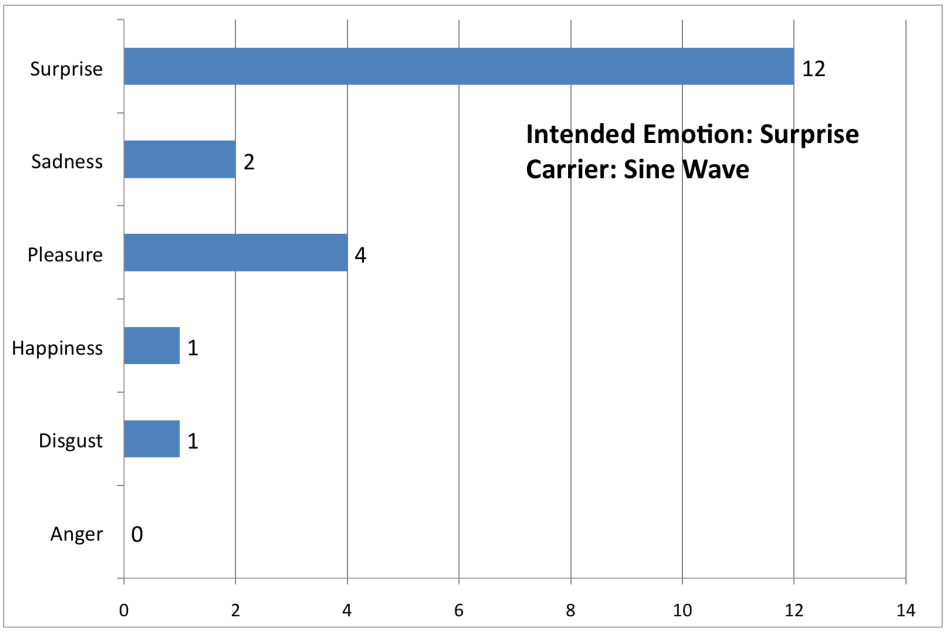


Figure A.20: *Participant's ratings for surprise based on sine wave*

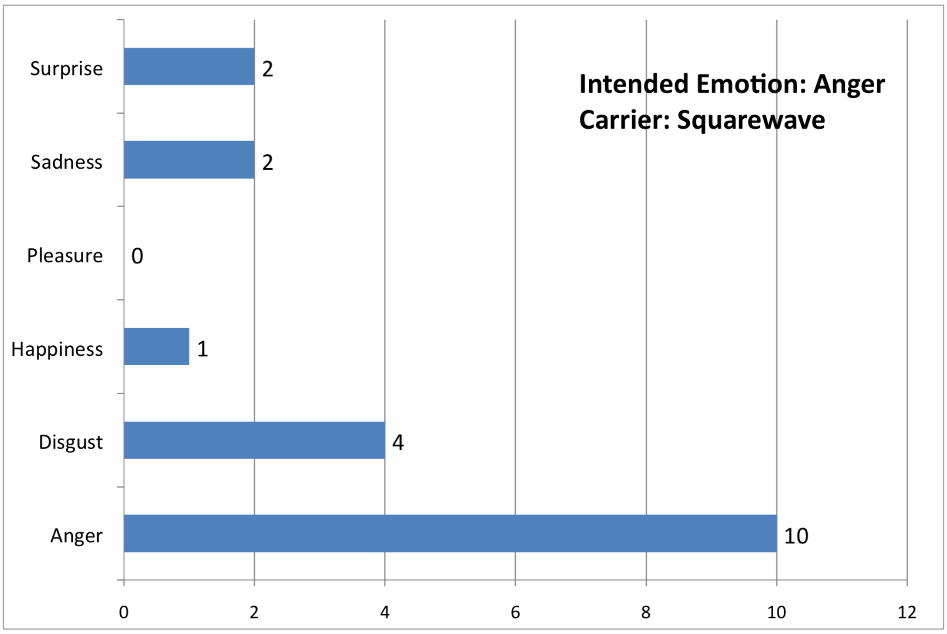


Figure A.21: *Participant's ratings for anger based on square wave*

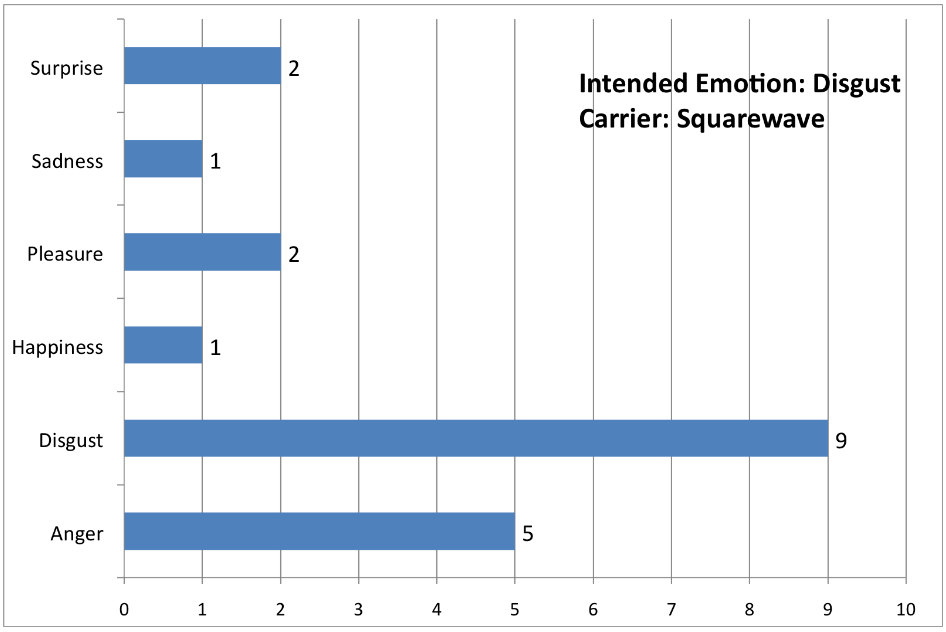


***Figure A.22**: Participant's ratings for disgust based on square wave*

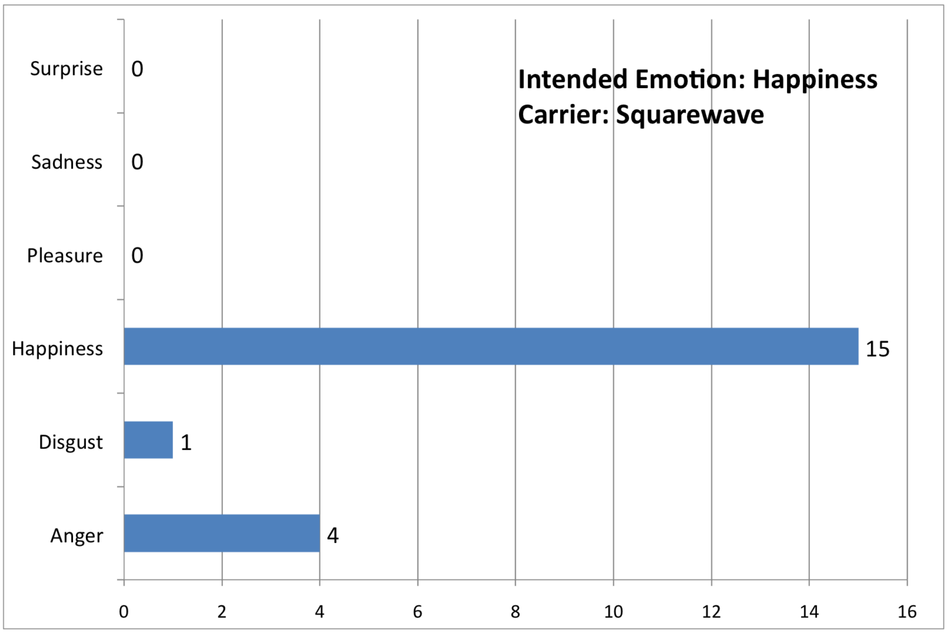


***Figure A.23**: Participant's ratings for happiness based on square wave*

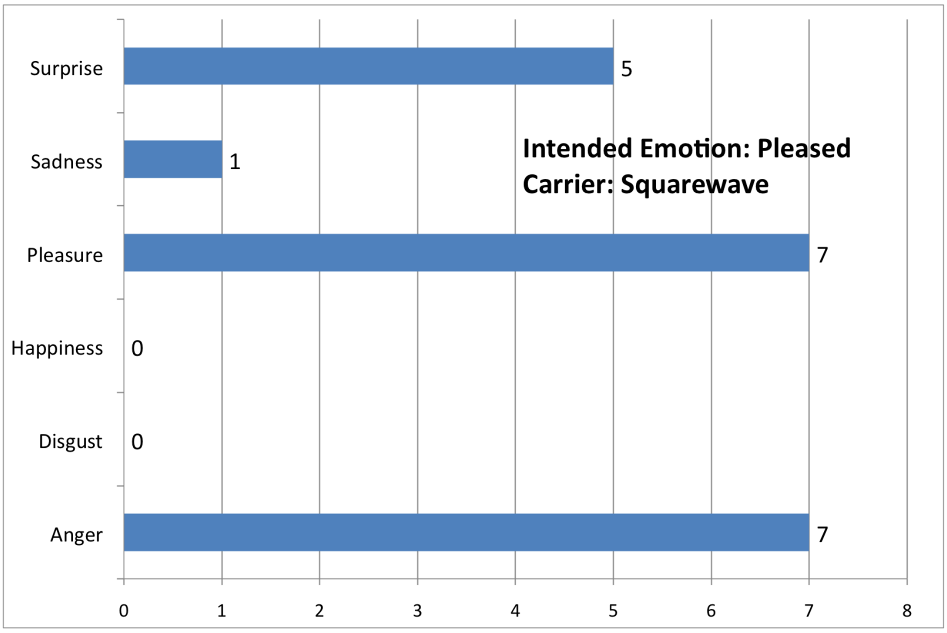


Figure A.24: *Participant's ratings for pleasure based on square wave*

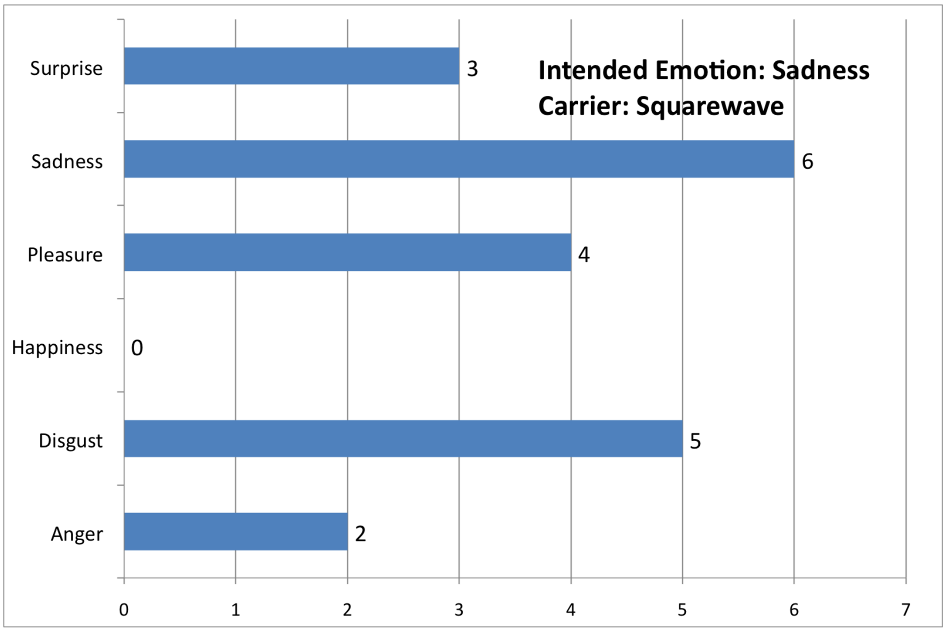


Figure A.25: *Participant's ratings for sadness based on square wave*

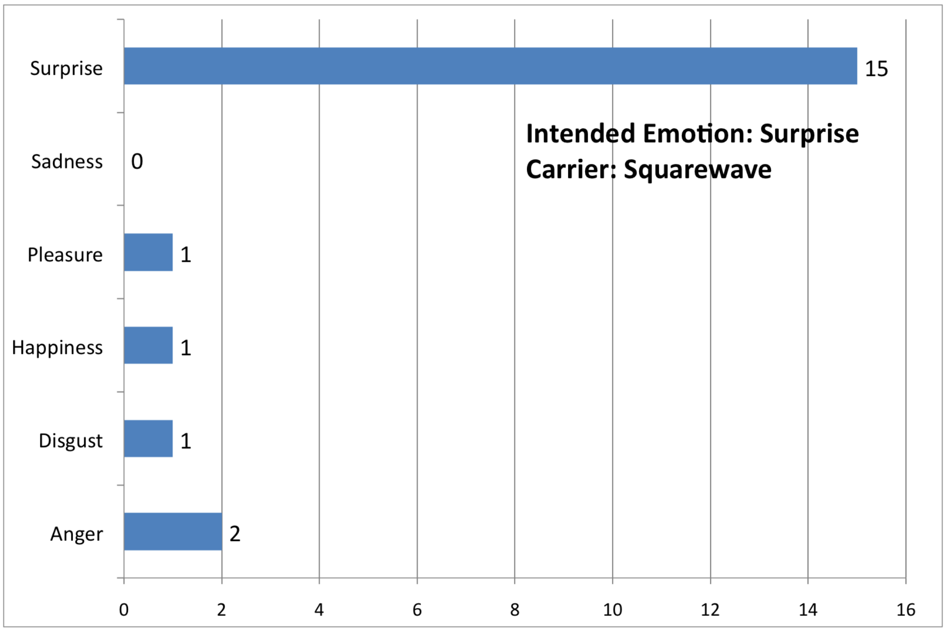


Figure A.26: *Participant's ratings for surprise based on square wave*

A.2 Phrasing Recommendations for Expressing Personality

In this section we will list all recommendations for phrasing and modelling responses of talking objects. As outlined in Section 6.3, each of these recommendations belongs to one of the categories *Dialogue Behaviour* (DB), *Speaking Style* (SS) and *Lexical Choice* (LC).

A.2.1 Extraversion

Recommendations for expressing a **high** degree of **extraversion**:

- Immediately greets the user upon contact, such as touching it (DB)
- Comparably more elaborate replies (DB)
- Frequent use of terms from a social context or describing positive emotions (SS)
- Usage of colloquial phrases, based on verbs, adverbs and pronouns (SS)
- Avoidance of *maybe, perhaps* and extensive usage of numbers (LC)
- Preferred bigrams: *a bit, a couple, other than, able to, want to, looking forward* and similar ones (LC)

Recommendations for expressing a **low** degree of **extraversion**:

- Only speaks when directly addressed by the user (DB)
- Replies with relatively short sentences (DB)
- Style of speech is more formal, based on adjectives and subjectives (SS)
- Prefers negative references of self, such as ”I don’t” (SS)
- Preferred bigrams are ”trying to”, ”going to”, ”should be” and similar ones (LC)
- Uses description of quantities such as ”loads of”, ”all the” and others (LC)
- Greets using ”Hello” or ”Good day” (LC)

A.2.2 Agreeableness

Recommendations for expressing a **high** degree of **agreeableness**:

- Praises and pays compliments (”A very good choice”) (DB)

- A generally friendly attitude (see also recommendations for extraversion) (SS)
- Expresses politeness (”please”, ”thank you”, ”Goodbye”) (SS)
- Avoids terms associated with insecurity, e.g. ”should”, ”would”, ”perhaps” and ”maybe” (LC)
- Applies *inclusives* and avoids *exclusives* (LC)

Recommendations for expressing a **low** degree of **agreeableness**:

- Replies generally unfriendly and huffish (SS)
- Prefers *exclusives* (e.g. ”without”, ”except”) over *inclusives* (LC)
- Applies terms that suggest anger (”hate”, ”angry”) (LC)

A.2.3 Conscientiousness

Recommendations for expressing a **high** degree of **conscientiousness**:

- Uses longer words, but replies relatively short in general (DB)
- Uses adjectives and adverbs less often (SS)
- Uses negations less often (SS)
- Exposes a high degree of determination by avoiding terms such as ”should”,”could” or ”would” (LC)
- Avoids the anticipatory tense (”about to”) (LC)

Recommendations for expressing a **low** degree of **conscientiousness**:

- Avoids negations, particularly if they relate to inabilities of the speaker (SS)
- Uses dependant clauses, often started with terms such as ”although” and ”whenever” (SS)
- Applies filler words such as ”you know” (LC)
- Uses relatively short words (LC)
- Often applies *discrepancies* (e.g. ”should”, ”could”) (LC)

A.2.4 Openness

Recommendations for expressing a **high** degree of **openness**:

- Seldom uses present tense and first person singular (SS)
- Uses longer and uncommon terms (LC)
- Prefers terms of the LIWC categories *communication*, *insight* and *tentativity* (LC)
- Increased use of articles (LC)

Recommendations for expressing a **low** degree of **openness**:

- Prefers present tense (SS)
- Often uses self references (”me”, ”my”, ”I”) (SS)
- Seldom uses words with more than 6 characters (LC)
- Avoids terms of the categories *tentativity* and *insight* (LC)

List of Figures

List of Tables

References

[Aaker, 1997] Aaker, J. (1997). Dimensions of brand personality. *Journal of Marketing research*, 34(3):347–356.

[Aarts and Marzano, 2003] Aarts, E. and Marzano, S. (2003). *The New Everyday: Visions of Ambient Intelligence*. Rotterdam, The Netherlands: 010 Publishing.

[Abrams, 1999] Abrams, R. (1999). Adventures in Tangible Computing: The Work of Interaction designer 'Durrell Bishop' In Context. Master's thesis, Royal College of Art, London, UK.

[Aish and Noakes, 1984] Aish, R. and Noakes, P. (1984). Architecture without numbers. *Computer Aided Design*, 16(6):321–328.

[André et al., 2000] André, E., Klesen, M., Gebhard, P., Allen, S., and Rist, T. (2000). Integrating models of personality and emotions into lifelike characters. *Affective interactions: Towards a New Generation of Computer Interfaces*, LNCS 1814:150–165.

[André et al., 1998] André, E., Rist, T., and Müller, J. (1998). Webpersona: A life-like presentation agent for the world-wide web. *Knowledge-Based Systems*, 11(1):25–36.

[Anselmi and Zemanek, 1997] Anselmi, K. and Zemanek, J. (1997). Relationship selling: how personal characteristics of salespeople affect buyer satisfaction. *Journal of Social Behavior and Personality*, 12:539–550.

[Apple et al., 1979] Apple, W., Streeter, L., and Krauss, R. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37:715–727.

[Bargar et al., 1994] Bargar, R., Choi, I., Das, S., and Goudeseune, C. (1994). Model-based interactive sound for an immersive virtual environment. In *Proceedings of the International Computer Music Conference*, pages 471–474. San Francisco, USA: International Computer Music Association.

[Bartels, 2004] Bartels, N. (2004). Pyramid Control. Bachelor's thesis, Saarland University, Saarbrücken, Germany.

[Beaudouin-Lafon, 2000] Beaudouin-Lafon, M. (2000). Instrumental interaction: an interaction model for designing post-wimp user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'06)*, pages 446–453. New York, NY, USA: ACM.

[Beigl et al., 2005] Beigl, M., Decker, C., Krohn, A., Riedel, T., and Zimmer, T. (2005). μParts: Low Cost Sensor Networks at Scale. In *Adjunct Proceedings of the International Conference on Ubiquitous Computing (UbiComp'05)*. New York, NY, USA: ACM.

[Beigl and Gellersen, 2003] Beigl, M. and Gellersen, H. (2003). Smart-Its: An Embedded Platform for Smart Objects. In *Proceedings of the Smart Objects Conference (SOC'03)*, pages 15–17. Grenoble, France: University Press.

[Beigl et al., 2001] Beigl, M., Gellersen, H., and Schmidt, A. (2001). Mediacups: Experience with Design and Use of Computer-Augmented Everyday Artefacts. *Computer Networks*, 35(4):401–409.

[Belin et al., 2008] Belin, P., Fillion-Bilodeau, S., and Gosselin, F. (2008). The montreal affective voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, 40(2):531–539.

[Benford et al., 2005] Benford, S., Schnädelbach, H., Koleva, B., Anastasi, R., Greenhalgh, C., Rodden, T., Green, J., Ghali, A., Pridmore, T., Gaver, B., Boucher, A., Walker, B., Pennington, S., Schmidt, A., Gellersen, H., and Steed, A. (2005). Expected, sensed, and desired: A framework for designing sensing-based interaction. *ACM Transactions on Computer-Human Interaction*, 12(1):3–30.

[Berscheid and Reis, 1998] Berscheid, E. and Reis, H. (1998). Attraction and close relationships. *The handbook of social psychology*, 2:193–281.

[Berwanger, 2004] Berwanger, M. (2004). Smart Clip. Bachelor's thesis, Saarland University, Saarbrücken, Germany.

[Bickmore et al., 2005] Bickmore, T. W., Caruso, L., and Clough-Gorr, K. (2005). Acceptance and usability of a relational agent interface by urban older adults. In *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems (CHI'05)*, pages 1212–1215. New York, NY, USA: ACM.

[Bickmore and Picard, 2005] Bickmore, T. W. and Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 12(2):293–327.

[Bohn et al., 2004] Bohn, J., Coroama, V., Langheinrich, M., Mattern, F., and Rohs, M. (2004). Living in a world of smart everyday objects – social, economic, and ethical implications. *Journal of Human and Ecological Risk Assessment*, 10(5):763–786.

[Borchers et al., 2002] Borchers, J., Ringel, M., Tyler, J., and Fox, A. (2002). Stanford interactive workspaces: a framework for physical and graphical user interface prototyping. *IEEE Wireless Communications*, 9(6):64–69.

[Bradley and Lang, 2000] Bradley, M. and Lang, P. (2000). Affective reactions to acoustic stimuli. *Psychophysiology*, 37:204–215.

[Bradski, 1998] Bradski, G. R. (1998). Computer Vision Face Tracking For Use in a Perceptual User Interface. In *Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision (WACV'98)*, pages 214–219. Los Alamitos, CA, USA: IEEE Computer Society Press.

[Brandherm et al., 2010] Brandherm, B., Haupert, J., Kröner, A., Schmitz, M., and Lehmann, F. (2010). Roles and rights management concept with identification by electronic identity card. In *Proceedings of the 8th Annual IEEE International Conference on Pervasive Computing and Communications (PerCom 2010)*, pages 768–771, Germany. Los Alamitos, CA, USA: IEEE Computer Society Press. ISBN 978-1-4244-5328-3.

[Brandtzaeg et al., 2004] Brandtzaeg, Bae, P., Følstad, A., and Heim, J. (2004). *Enjoyment: lessons from Karasek*, pages 55–65. Norwell, MA, USA: Kluwer Academic Publishers.

[Breazeal et al., 2004] Breazeal, C., Hoffman, G., and Lockerd, A. (2004). Teaching and Working with Robots as a Collaboration. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '04)*, pages 1030–1037. Los Alamitos, CA, USA: IEEE Computer Society Press.

[Broll et al., 2009] Broll, G., Rukzio, E., Paolucci, M., Wagner, M., Schmidt, A., and Hussmann, H. (2009). PERCI: Pervasive Service Interaction with the Internet of Things. *IEEE Internet Computing*, 6(1):74–81.

[Brown et al., 1975] Brown, B., Strong, W., and Rencher, A. (1975). Acoustic determinants of the perceptions of personality from speech. *International Journal of the Sociology of Language*, 6:11–32.

[Brubaker et al., 2007] Brubaker, M. A., Fleet, D. J., and Hertzmann, A. (2007). Physics-Based Person Tracking Using Simplified Lower-Body Dynamics. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–8. Los Alamitos, CA, USA: IEEE Computer Society Press.

[Brunette et al., 2003] Brunette, W., Hartung, C., Nordstrom, B., and Borriello, G. (2003). Proximity interactions between wireless sensors and their application. In *Proceedings of the 2nd International Conference on Wireless Sensor Networks and Applications (WSNA '03)*, pages 30–37. New York, NY, USA: ACM.

[Bullock, 1985] Bullock, M. (1985). Animism in Childhood Thinking: A New Look at an Old Question. *Developmental Psychology*, 21(2):217–25.

[Burneleit et al., 2009] Burneleit, E., Hemmert, F., and Wettach, R. (2009). Living interfaces: the impatient toaster. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction (TEI'09)*, pages 21–22. New York, NY, USA: ACM.

[Butz et al., 2004] Butz, A., Groß, M., and Krüger, A. (2004). Tuister: a tangible ui for hierarchical structures. In *Proceedings of the 9th International Conference on Intelligent User Interfaces (IUI '04)*, pages 223–225. New York, NY, USA: ACM.

[Butz and Krüger, 2006] Butz, A. and Krüger, A. (2006). A mixed reality room following the generalized peephole metaphor. *IEEE Computer Graphics & Applications*, 26(1):56–63.

[Butz and Schmitz, 2005] Butz, A. and Schmitz, M. (2005). Design and Applications of a Beer Mat for Pub Interaction. In *Proceedings of the Seventh International Conference on Ubiquitous Computing (UbiComp'05)*. New York, NY, USA: ACM.

[Butz et al., 2005] Butz, A., Schmitz, M., Krüger, A., and Hullmann, H. (2005). Tangible UIs for Media Control: Probes into the Design Space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'05)*, pages 957–971. New York, NY, USA: ACM.

[Byrne and Griffitt, 1969] Byrne, D. and Griffitt, W. (1969). Similarity and awareness of similarity of personality characteristics determinants of attraction. *Journal of Experimental Research in Personality*, 3:179–186.

[Caporael and Heyes, 1997] Caporael, L. and Heyes, C. (1997). *Anthropomorphism, Anecdotes, and Animals*, chapter Why anthropomorphize? Folk psychology and other stories, pages 59–73. Albany, NY, USA: State University of New York Press.

[Carsten, 2009] Carsten, S. (2009). Multispektrale Vor-Ort-Evaluation des VirtualConstructor-Exponats der VW-Autostadt Wolfsburg. Diplomarbeit, Technische Universität Ilmenau, Germany.

[Cassell et al., 1998] Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Villhjálmsson, H., and Yan, H. (1998). An architecture for embodied conversational characters. In *Proceedings of the First Workshop on Embodied Conversational Characters*, pages 109–120, Tahoe City, California.

[Cassell et al., 1999] Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H., and Yan, H. (1999). Conversation as a system framework: Designing embodied conversational agents. *Embodied Conversational Agents*, pages 29–63.

[Cheok et al., 2008] Cheok, A. D., Kok, R. T., Tan, C., Newton Fernando, O. N., Merritt, T., and Sen, J. Y. P. (2008). Empathetic living media. In *Proceedings of the 7th ACM conference on Designing interactive systems (DIS'08)*, pages 465–473. New York, NY, USA: ACM.

[Chi et al., 2007] Chi, P.-y., Chen, J.-h., Liu, S.-y., and Chu, H.-h. (2007). Designing Smart Living Objects - Enhancing vs. Distracting Traditional Human-Object Interaction. In *Proceedings of the International Conference on Human-Computer Interaction (HCII'07)*, LNCS 4551, pages 788–797. Berlin, Heidelberg, Germany: Springer.

[Choi and Yoo, 2008] Choi, J. and Yoo, C.-W. (2008). Connect with Things through Instant Messaging. In Floerkemeier, C., Langheinrich, M., Fleisch, E., Mattern, F., and Sarma, S. E., editors, *Proceedings of the First International Conference on The Internet of Things (IOT 2008)*, LNCS 4952, pages 276–288. Berlin, Heidelberg, Germany: Springer.

[Coen, 1998] Coen, M. H. (1998). Design principles for intelligent environments. In *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence (AAAI '98/IAAI '98)*, pages 547–554. Menlo Park, CA, USA: American Association for Artificial Intelligence.

[Cohen and McGee, 2004] Cohen, P. and McGee, D. (2004). Tangible multimodal interfaces for safety-critical applications. *Communications of the ACM*, 47(1):41–46.

[Comption and Jansen, 1990] Comption, P. and Jansen, R. (1990). A philosophical basis for knowledge acquisition. *Knowledge Acquisition*, 2:241–257.

[Costa and McCrae, 1985] Costa, P. and McCrae, R. (1985). *The NEO personality inventory manual*. Odessa, FL, USA: Psychological Assessment Resources.

[Crutzen, 2000] Crutzen, C. (2000). *Interactie, een wereld van verschillen. Een visie op informatica vanuit genderstudies.* PhD thesis, Open Universiteit Nederland, Heerlen, The Netherlands.

[Cucchiara et al., 2004] Cucchiara, R., Grana, C., Tardini, G., and Vezzani, R. (2004). Probabilistic People Tracking for Occlusion Handling. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004)*, pages 132–135. Los Alamitos, CA, USA: IEEE Computer Society Press.

[Cucchiara et al., 2005] Cucchiara, R., Prati, A., Benini, L., and Farella, E. (2005). T_PARK: Ambient Intelligence for Security in Public Parks. In *Proceedings of IEE International Conference on Intelligent Environments*, pages 243–251. Amsterdam, The Netherlands: IOS Press.

[Damon and Phelps, 1989] Damon, W. and Phelps, E. (1989). *Peer relationships in child development*, chapter Strategic uses of peer learning in children's education, pages 135–157. Hoboken, NJ, USA: Wiley.

[Delaney and Dobson, 2009] Delaney, K. and Dobson, S. (2009). Augmenting Materials to Build Cooperating Objects. *Ambient Intelligence with Microsystems*, 18(1):19–46.

[Djajadiningrat et al., 2000] Djajadiningrat, J. P., Overbeeke, C. J., and Wensveen, S. A. G. (2000). Augmenting fun and beauty: a pamphlet. In *Proceedings of the Internationa Conference on Designing Augmented Reality Environments (DARE'00)*, pages 131–134. New York, NY, USA: ACM.

[Djajadiningrat et al., 2002] Djajadiningrat, T., Overbeeke, K., and Wensveen, S. (2002). But how, Donald, tell us how? On the creation of meaning in interaction design through feedforward and inherent feedback. In *Proceedings of the 4th conference on designing* *interactive systems: processes, practices, methods, and techniques*, pages 285–291. New York, NY, USA: ACM.

[Dourish, 2001] Dourish, P. (2001). *Where the Action Is. The Foundations of Embodied Interaction*. Bradford Book, MIT Press.

[Ducatel et al., 2001] Ducatel, K., Bogdanowicz, M., Scapolo, F., Leijten, J., and Burgelma, J. C. (2001). Scenarios for Ambient Intelligence in 2001 (ISTAG 2001 Final Report). Technical report, IST Advisory Group.

[Duffy, 2003] Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3-4):177 – 190.

[Duggan and Deegan, 2003] Duggan, B. and Deegan, M. (2003). Considerations in the usage of text to speech (tts) in the creation of natural sounding voice enabled web systems. In *Proceedings of the 1st International Symposium on Information and communication technologies (ISICT '03)*, pages 433–438. Dublin, Ireland: Trinity College Dublin.

[Dutoit, 1997] Dutoit, T. (1997). *An Introduction to Text-To-Speech Synthesis*. Berlin, Heidelberg, Germany: Springer.

[Edworthy and Hellier, 2006] Edworthy, J. and Hellier, E. (2006). Alarms and human behaviour: implications for medical alarms. *BJA: British Journal of Anaesthesia*, 97(1):12–17.

[Ekman, 1993] Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, 48(4):384–392.

[Enrico Rukzio, 2007] Enrico Rukzio, Gregor Broll, K. L. A. S. (2007). Mobile Interaction with the Real World: An Evaluation and Comparison of Physical Mobile Interaction Techniques. In *Ambient Intelligence*, volume LNCS 4794, pages 1–18. Berlin, Heidelberg, Germany: Springer.

[Enrico Rukzio, 2008] Enrico Rukzio, Gregor Broll, S. W. (2008). The physical mobile interaction framework (pmif). Technical report lmu-mi-2008-2, Munich, Germany: Ludwig-Maximilians-Universität München.

[Ess et al., 2008] Ess, A., Leibe, B., Schindler, K., and Van Gool, L. (2008). A Mobile Vision System for Robust Multi-Person Tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, pages 1–8. Los Alamitos, CA, USA: IEEE Computer Society Press.

[Estrin et al., 2002] Estrin, D., Culler, D., Pister, K., and Sukhatme, G. (2002). Connecting the physical world with pervasive networks. *IEEE Pervasive Computing*, 1(1):59–69.

[Feld et al., 2010] Feld, M., Schwartz, T., and Müller, C. (2010). This Is Me: Using Ambient Voice Patterns for In-Car Positioning. *Proceedings of the First International Joint Conference on Ambient Intelligence*, pages 290–294.

[Fernaeus et al., 2010] Fernaeus, Y., Håkansson, M., Jacobsson, M., and Ljungblad, S. (2010). How do you play with a robotic toy animal? A long-term study of Pleo. In *Proceedings of the 9th International Conference on Interaction Design and Children (IDC '10)*, pages 39–48. New York, NY, USA: ACM.

[Fishkin et al., 2000a] Fishkin, K., Goldberg, D., Gujar, A., Harrison, B., Mynatt, E., Stone, M., and Want, R. (2000a). Zoomorphic computer user interface. US Patent 6,160,540.

[Fishkin, 2004] Fishkin, K. P. (2004). A taxonomy for and analysis of tangible interfaces. *Personal Ubiquitous Computing*, 8(5):347–358.

[Fishkin et al., 2000b] Fishkin, K. P., Gujar, A., Harrison, B. L., Moran, T. P., and Want, R. (2000b). Embodied user interfaces for really direct manipulation. *Communications of the ACM*, 43(9):74–80.

[Fiske and Taylor, 1991] Fiske, S. and Taylor, S. (1991). *Social cognition*. New York, NY, USA: McGraw-Hill.

[Fitzmaurice et al., 1995] Fitzmaurice, G., Ishii, H., and Buxton, W. (1995). Bricks: Laying the Foundations for Graspable User Interfaces. In *Proceedings of Conference on Human Factors in Computing Systems (CHI'95)*, pages 442–449. New York, NY, USA: ACM.

[Fitzmaurice, 1996] Fitzmaurice, G. W. (1996). *Graspable User Interfaces*. PhD thesis, University of Toronto, Canada.

[Frazer et al., 1980] Frazer, J., Frazer, J., and Frazer, P. (1980). Intelligent physical three-dimensional modelling systems. In *Proceedings of the International Conference on Computer Graphics*, pages 359–370. Los Alamitos, CA, USA: IEEE Computer Society Press.

[Frazer et al., 1982] Frazer, J., Frazer, J., and Frazer, P. (1982). Three-dimensional data input devices. In *Proceedings of Computer Graphics in the Building Process*, pages 409–416. Washington, DC, USA: National Academy of Sciences.

[Freeman et al., 1999] Freeman, E., Arnold, K., and Hupfer, S. (1999). *JavaSpaces principles, patterns, and practice*. Addison-Wesley Longman Ltd. Essex, UK, UK.

[Friedman et al., 2003] Friedman, B., Kahn Jr, P., and Hagman, J. (2003). Hardware companions?: What online AIBO discussion forums reveal about the human-robotic relationship. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'03)*, pages 273–280. New York, NY, USA: ACM.

[Froumentin, 2004] Froumentin, M. (2004). Extensible multimodal annotation markup language (EMMA): Invited talk. In *Proceeedings of the Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology (NLPXML '04)*, pages 33–33. Morristown, NJ, USA: Association for Computational Linguistics.

[Fujinami et al., 2005] Fujinami, K., Kawsar, F., and Nakajima, T. (2005). *Pervasive Computing*, chapter AwareMirror: A Personalized Display Using a Mirror, pages 315–332. Berlin, Heidelberg, Germany: Springer.

[Gebhard, 2005] Gebhard, P. (2005). ALMA: A Layered Model of Affect. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 29–36. New York, NY, USA: ACM.

[Gebhard, 2009] Gebhard, P. (2009). *Emotionalisierung interaktiver Virtueller Charaktere - Ein mehrschichtiges Computermodell zur Erzeugung und Simulation von Gefühlen in Echtzeit*, volume 315 of *DISKI*. Berlin, Germany: Akademische Verlagsgesellschaft Aka.

[Gebhard et al., 2003] Gebhard, P., Kipp, M., Klesen, M., and Rist, T. (2003). Authoring scenes for adaptive, interactive performances. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 725–732. New York, NY, USA: ACM.

[Gebhard et al., 2008] Gebhard, P., Schröder, M., Charfuelan, M., Endres, C., Kipp, M., Pammi, S., Rumpler, M., and Türk, O. (2008). IDEAS4Games: building expressive virtual characters for computer games. In *Proceedings of the 8th International Conference on Intelligent Virtual Agents (IVA 2008)*, pages 426–440. Berlin, Heidelberg, Germany: Springer.

[Gibson, 1977] Gibson, J. J. (1977). *Perceiving, Acting, and Knowing. Toward an ecological psychology*, chapter The Theory of Affordances, pages 67–82. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

[Gilbertson et al., 1998] Gilbertson, J., Dindia, K., and Allen, M. (1998). Relational continuity constructional units and the maintenance of relationships. *Journal of Social and Personal Relationships*, 15(6):774.

[Gill et al., 2006] Gill, A. J., Oberlander, J., and Austin, E. (2006). The perception of e-mail personality at zero-acquaintance. *Personality and Individual Differences*, 40:497–507.

[Göbel et al., 2004] Göbel, S., Schneider, O., Iurgel, I., Feix, A., Knöpfle, C., and Rettig, E. (2004). Virtual Human: Storytelling and computer graphics for a virtual human platform. In *Proceedings of Technologies for Interactive Digital Storytelling and Entertainment 2004*, LNCS 3105, pages 79–88. Berlin, Heidelberg, Germany: Springer.

[Gutek et al., 2000] Gutek, B., Cherry, B., Bhappu, A., Schneider, S., and Woolf, L. (2000). Features of service relationships and encounters. *Work and Occupations*, 27(3):319.

[Guthrie, 1997] Guthrie, S. (1997). *Anthropomorphism, Anecdotes, and Animals*, chapter Anthropomorphism: A definition and a theory, pages 50–58. Albany, NY, USA: State University of New York Press.

[Guthrie and MyiLibrary, 1993] Guthrie, S. and MyiLibrary (1993). *Faces in the clouds: A new theory of religion*. Oxford, UK: Oxford University Press.

[Hager, 1987] Hager, W. (1987). *Allgemeine experimentelle Psychologie*, chapter Grundlagen einer Versuchsplanung zur Prüfung empirischer Hypothesen in der Psychologie, pages 43–264. Jena, Germany: Gustav Fischer Verlag, Stuttgart.

[Haritaoglu et al., 1998] Haritaoglu, I., Harwood, D., and Davis, L. S. (1998). W4S: A real-time system for detecting and tracking people in 2 1/2-D. In *Proceedings of the European Conference on Computer Vision*, pages 877–892. Berlin, Heidelberg, Germany: Springer.

[Hartley and Zisserman, 2004] Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge, UK: Cambridge University Press.

[Hassenzahl et al., 2001] Hassenzahl, M., Beu, A., and Burmester, M. (2001). Engineering joy. *IEEE Software*, 1(2):70–76.

[Hayes-Roth and Doyle, 1998] Hayes-Roth, B. and Doyle, P. (1998). Animate characters. *Autonomous Agents and Multi-Agent Systems*, 1(2):195–230.

[Hemmert, 2008] Hemmert, F. (2008). Ambient Life: Permanent Tactile Life-like Actuation as a Status Display in Mobile Phones. In *Adjunct Proceedings of the 21st annual ACM symposium on User Interface Software and Technology (UIST'08)*. New York, NY, USA: ACM.

[Hemmert et al., 2010] Hemmert, F., Hamann, S., Löwe, M., Wohlauf, A., and Joost, G. (2010). Shape-changing mobiles: tapering in one-dimensional deformational displays in mobile phones. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction (TEI '10)*, pages 249–252. New York, NY, USA: ACM.

[Hemmert and Joost, 2008] Hemmert, F. and Joost, G. (2008). Ambient life: Calm and excited permanent tactile life-like actuation as a status display in mobile phones. In *Proceedings of the Conference on Design and Emotion*, Hong Kong.

[Hill and Culler, 2002] Hill, J. and Culler, D. (2002). Mica: A wireless platform for deeply embedded networks. *IEEE Micro*, 22(6):12–24.

[Hoffman and Breazeal, 2007] Hoffman, G. and Breazeal, C. (2007). Effects of Anticipatory Action on Human-Robot Teamwork Efficiency, Fluency, and Perception of Team. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI'08)*, pages 1–8. New York, NY, USA: ACM.

[Hollinger, 2007] Hollinger, B. (2007). Dialogue Shell for Talking Objects. Bachelor's thesis, Saarland University, Saarbrücken, Germany.

[Holman and Vertegaal, 2008] Holman, D. and Vertegaal, R. (2008). Organic user interfaces: designing computers in any way, shape, or form. *Communications of the ACM*, 51(6):48–55.

[Holmquist et al., 2004a] Holmquist, L. E., Gellersen, H.-W., Kortuem, G., Schmidt, A., Strohbach, M., Antifakos, S., Michahelles, F., Schiele, B., Beigl, M., and Mazé, R. (2004a). Building Intelligent Environments with Smart-Its. *IEEE Computer Graphics & Applications*, 24(1):56–64.

[Holmquist et al., 2004b] Holmquist, L. E., Schmidt, A., and Ullmer, B. (2004b). Tangible interfaces in perspective: Guest editors' introduction. *Personal Ubiquitous Computing*, 8(5):291–293.

[Hong et al., 2008] Hong, D., Höllerer, T., Haller, M., Takemura, H., Cheok, A. D., Kim, G. J., Billinghurst, M., Woo, W., Hornecker, E., Jacob, R. J. K., Hummels, C., Ullmer, B., Schmidt, A., van den Hoven, E., and Mazalek, A. (2008). Advances in tangible interaction and ubiquitous virtual reality. *IEEE Pervasive Computing*, 7(2):90–96.

[Hornecker, 2004] Hornecker, E. (2004). *Tangible User Interfaces als kooperationsunterstützendes Medium*. PhD thesis, Universität Bremen, Germany.

[Hornecker and Bruns, 2004] Hornecker, E. and Bruns, F. (2004). Interaktion im Sensoric Garden. In *Tagungsband der Mensch und Computer (MuC 2004)*, pages 65–74. Munich, Germany: Oldenbourg Verlag.

[Hornecker and Buur, 2006] Hornecker, E. and Buur, J. (2006). Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'06)*, pages 437–446. New York, NY, USA: ACM.

[Horvitz et al., 1998] Horvitz, E., Breese, J., Heckerman, D., Hovel, D., and Rommelse, K. (1998). The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 256–265. San Francisco, CA, USA: Morgan Kaufmann.

[Hurtienne and Israel, 2007] Hurtienne, J. and Israel, J. H. (2007). Image schemas and their metaphorical extensions: intuitive patterns for tangible interaction. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction (TEI'07)*, pages 127–134. New York, NY, USA: ACM.

[Inoue et al., 2008] Inoue, K., Wada, K., and Ito, Y. (2008). Effective Application of Paro: Seal Type Robots for Disabled People in According to Ideas of Occupational Therapists. In *Proceedings of the International Conference on Computers Helping People with Special Needs (ICCHP 2008)*, pages 1321–1324. Berlin, Heidelberg, Germany: Springer.

[Ishii et al., 1999] Ishii, H., Fletcher, H., Lee, J., Choo, S., Berzowska, J., Wisneski, C., Cano, C., Hernandez, A., and Bulthaup, C. (1999). MusicBottles. In *Proceedings of the International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'99)*, page 174. New York, NY, USA: ACM.

[Ishii et al., 2002] Ishii, H., J., U., Chak, D., Piper, B., Ben-Joseph, E., Yeung, L., and Kanji, Z. (2002). Augmented Urban Planning Workbench: Overlaying Drawings, Physical Models and Digital Simulation. In *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 203–211. New York, NY, USA: ACM.

[Ishii et al., 2001] Ishii, H., Mazalek, A., and Lee, J. (2001). Bottles as a minimal interface to access digital information. In *Extended abstracts of the SIGCHI Conference on Human Factors in Computing Systems (CHI'01)*, pages 187–188. New York, NY, USA: ACM.

[Ishii and Ullmer, 1997] Ishii, H. and Ullmer, B. (1997). Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'97)*, pages 234–241. New York, NY, USA: ACM.

[ISO/IEC, 1996] ISO/IEC (1996). ISO 9241-10 Ergonomic requirements for office work with visual display terminals (VDTs) – Part 10: Dialogue principles. Technical report, International Organization for Standardization.

[ISO/IEC, 1998] ISO/IEC (1998). ISO 9241-11 Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. Technical report, International Organization for Standardization.

[Jacob et al., 2008] Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., and Zigelbaum, J. (2008). Reality-Based Interaction: A Framework for Post-WIMP Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'03)*, pages 201–210. New York, NY, USA: ACM.

[Jakobson, 1960] Jakobson, R. (1960). Linguistics and poetics. *Style in language*, pages 350–377.

[J.J. Bargh and Burrows, 1996] J.J. Bargh, M. C. and Burrows, L. (1996). Automaticity of social behaviour: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71:230–244.

[Johanson and Fox, 2002] Johanson, B. and Fox, A. (2002). The event heap: A coordination infrastructure for interactive workspaces. In *Proceedings of the Fourth IEEE Workshop on Mobile Computing Systems and Applications*, pages 83–93. Los Alamitos, CA, USA: IEEE Computer Society Press.

[Jones and Marsden, 2006] Jones, M. and Marsden, G. (2006). *Mobile Interaction Design*. Hoboken, NJ, USA: Wiley.

[Jordà et al., 2007] Jordà, S., Geiger, G., Alonso, M., and Kaltenbrunner, M. (2007). The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st International Conference on Tangible and Embedded interaction (TEI'07)*, pages 139–146. New York, NY, USA: ACM.

[Jordà et al., 2005] Jordà, S., Kaltenbrunner, M., Geiger, G., and Bencina, R. (2005). The Reactable. In *Proceedings of the International Computer Music Conference (ICMC 2005)*, pages 579–582. San Francisco, CA, USA: International Computer Music Association.

[Jung, 2009] Jung, R. (2009). *Ambiente Audionotifikation - Ein System zur kontext-sensitiven Integration von nicht-intrusiven Notifikationssignalen in emotionsklassifizierten Soundscapes*. PhD thesis, Saarland University, Saarbrücken, Germany.

[Kain and Macon, 2002] Kain, A. and Macon, M. (2002). Design and evaluation of a voice conversion algorithm based on spectral envelope mapping and residual prediction. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'01)*, volume 2, pages 813–816. Los Alamitos, CA, USA: IEEE Computer Society Press.

[Kaiser, 2004] Kaiser, M. (2004). Smart Security Tag. Bachelor's thesis, Saarland University, Saarbrücken, Germany.

[Kawsar et al., 2005] Kawsar, F., Fujinami, K., and Nakajima, T. (2005). Augmenting everyday life with sentient artefacts. In *Proceedings of the 2005 joint conference on Smart Objects and Ambient Intelligence (SOC-EUSAI'05)*, pages 141–146. New York, NY, USA: ACM.

[Kestner et al., 2009] Kestner, J., Leithinger, D., Jung, J., and Petersen, M. (2009). Proverbial wallet: tangible interface for financial awareness. In *TEI '09: Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pages 55–56. New York, NY, USA: ACM.

[Kidd et al., 1999] Kidd, C., Orr, R., Abowd, G., Atkeson, C., Essa, I., MacIntyre, B., Mynatt, E., Starner, T., and Newstetter, W. (1999). The Aware Home: A Living Laboratory for Ubiquitous Computing Research. In *Proceedings of the Second International Workshop on Cooperative Buildings (CoBuild'99)*, pages 191–198. Berlin, Heidelberg, Germany: Springer.

[Kiesler, 1983] Kiesler, D. (1983). The 1982 Interpersonal Circle: A taxonomy for complementarity in human transactions. *Psychological Review*, 90(3):185–214.

[Kim et al., 2008] Kim, S., Kim, H., Lee, B., Nam, T.-J., and Lee, W. (2008). Inflatable Mouse: Volume-adjustable Mouse with Air-pressure-sensitive Input and Haptic Feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*, pages 211–224. New York, NY, USA: ACM.

[Klein et al., 2002] Klein, J., Moon, Y., and Picard, R. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14(2):119–140.

[Komatsu, 2006] Komatsu, T. (2006). Audio Subtle Expressions Affecting User's Perceptions. In *Proceedings of the 11th International Conference on Intelligent User Interfaces (IUI 2006)*, pages 306–308. New York, NY, USA: ACM.

[Komatsu and Yamada, 2008] Komatsu, T. and Yamada, S. (2008). Effect of Agent Appearance on People's Interpretation of Agent's Attitude. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*, pages 2919–2924. New York, NY, USA: ACM.

[Kortuem et al., 2007] Kortuem, G., Alford, D., Ball, L., Busby, J., Davies, N., Efstratiou, C., Finney, J., White, M., and Kinder, K. (2007). Sensor networks or smart artifacts? An exploration of organizational issues of an industrial health and safety monitoring system.

In *Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp 2007)*, LNCS 4717, pages 465–482. Berlin, Heidelberg, Germany: Springer.

[Krauss et al., 2002] Krauss, R., Freyberg, R., and Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology*, 38(6):618–625.

[Kröner et al., 2009] Kröner, A., Gebhard, P., Spassova, L., Kahl, G., and Schmitz, M. (2009). Informing Customers by Means of Digital Product Memories. In *Proceedings of the 1st International Workshop on Digital Object Memories. Workshop on Digital Object Memories (DOME-09), located at Intelligent Environments*, pages 21–26. Amsterdam, The Netherlands: IOS Press.

[Krüger et al., 2010] Krüger, A., Spassova, L., and Jung, R. (2010). Innovative retail laboratory - investigating future shopping technologies. *IT - information technology*, 52(2):114–118.

[Krumm et al., 2000] Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., and Shafer, S. A. (2000). Multi-Camera Multi-Person Tracking for EasyLiving. In *Workshop on Visual Surveillance*.

[Kruppa, 2005] Kruppa, M. (2005). *Migrating Characters: Effective User Guidance in Instrumented Environments*, volume 301. Berlin, Germany: Akademische Verlagsgesellschaft Aka.

[Kruppa et al., 2005] Kruppa, M., Spassova, L., and Schmitz, M. (2005). The Virtual Room Inhabitant–Intuitive Interaction with Intelligent Environments. In *Proceedings of the 18th Australian Joint Conference on Artificial Intelligence (AI05)*, LNCS 3809, pages 225–234. Berlin, Heidelberg, Germany: Springer.

[Larsson, 2010] Larsson, P. (2010). Tools for Designing Emotional Auditory Driver-Vehicle Interfaces. *Auditory Display*, LNCS 5954(1):1–11.

[Laukka et al., 2005] Laukka, P., Juslin, P., and Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion*, 19(5):633–653.

[Lee, 2008] Lee, E. (2008). Cyber physical systems: Design challenges. In *Object Oriented Real-Time Distributed Computing (ISORC), 2008 11th IEEE International Symposium on*, pages 363–369. IEEE.

[Lee and Nass, 2003] Lee, K. M. and Nass, C. (2003). Designing Social Presence of Social Actors in Human Computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'03)*, pages 289–296. New York, NY, USA: ACM.

[Lehman et al., 2001] Lehman, T., Cozzi, A., Xiong, Y., Gottschalk, J., Vasudevan, V., Landis, S., Davis, P., Khavar, B., and Bowman, P. (2001). Hitting the distributed computing sweet spot with TSpaces. *Computer Networks*, 35(4):457–472.

[Levinson, 1983] Levinson, S. (1983). *Pragmatics (Cambridge Textbooks in Linguistics)*. Cambridge, MA: Cambridge University Press.

[Lifton, 2007] Lifton, J. H. (2007). *Dual Reality: An Emerging Medium*. PhD thesis, School of Architecture and Planning, Massachusetts Institute of Technology (MIT), USA.

[Lombriser et al., 2009] Lombriser, C., Bulling, A., Breitenmoser, A., and Troster, G. (2009). Speech as a feedback modality for smart objects. *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*, pages 1–5.

[Löwgren, 2001] Löwgren, J. (2001). *Human Computer Interaction: Issues and Challenges*, chapter From HCI to Interaction Design, pages 29–43. Hershey, PA, USA: IGI Publishing.

[MacLean, 2000a] MacLean, K. (2000a). Designing with haptic feedback. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'00)*, pages 783–788. Los Alamitos, CA, USA: IEEE Computer Society Press.

[MacLean, 2000b] MacLean, K. (2000b). *Human and Machine Haptics*, chapter Application-Centered Haptic Interface Design. Cambridge, MA: MIT Press.

[Maes, 1994] Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM*, 37(7):30–40.

[Mattern, 2003] Mattern, F. (2003). From Smart Devices to Smart Everyday Objects. In *Proceedings of the Smart Objects Conference (SOC'03)*, pages 15–16. Grenoble, France: University Press.

[Mattern, 2004] Mattern, F. (2004). Ubiquitous Computing: Schlaue Alltagsgegenstände – Die Vision von der Informatisierung des Alltags. *Bulletin SEV/VSE*, 19:9–13.

[Mattern, 2005] Mattern, F. (2005). *E-Merging Media - Communication and the Media Economy of the Future*, chapter Ubiquitous Computing: Scenarios from an Informatised World, pages 145–163. Berlin, Heidelberg, Germany: Springer.

[McGee et al., 2000] McGee, D. R., Cohen, P. R., and Wu, L. (2000). Something from nothing: augmenting a paper-based work practice via multimodal interaction. In *Proceedings of the International Conference on Designing Augmented Reality Environments*, pages 71–80. New York, NY, USA: ACM.

[McGrenere and Ho, 2000] McGrenere, J. and Ho, W. (2000). Affordances: Clarifying and evolving a concept. In *Proceedings of the Graphics Interface conference (GI 2000)*, pages 179–186. New York, NY, USA: ACM.

[McKenna et al., 2000] McKenna, S. J., Jabri, S., Duric, Z., Rosenfeld, A., and Wechsler, H. (2000). Tracking Groups of People. *Computer Vision and Image Understanding*, 80(1):42–56.

[McRae and John, 1992] McRae, R. and John, O. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60:175–215.

[Mead, 1932] Mead, M. (1932). An Investigation of the Thought of Primitive Children, with Special Reference to Animism. *Journal of the Anthropological Institute of Great Britain and Ireland*, 62:173–190.

[Mereu and Kazman, 1996] Mereu, S. W. and Kazman, R. (1996). Audio Enhanced 3D Interfaces for Visually Impaired Users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'96)*, pages 72–78. New York, NY, USA: ACM.

[Minina et al., 2009] Minina, E., Schmitz, M., and Quraischy, H. (2009). Evaluating tangible interaction with real and virtual products – user feedback results. In *Proceedings of the IADIS International Conference on Interfaces and Human Computer Interaction (IHCI'09)*, pages 182–186. Lisbon, Portugal: IADIS press.

[Mithen, 1996] Mithen, S. (1996). *The prehistory of the mind: a search for the origins of art, religion and science*. London, UK: Thames and Hudson.

[Monk et al., 2002] Monk, A., Hassenzahl, M., Blythe, M., and D. Reed, D. (2002). Funology: designing enjoyment (workshop). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'02)*, pages 924–925. New York, NY, USA: ACM.

[Moon, 2000] Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research*, 26(4):323–339.

[Morkes et al., 1999] Morkes, J., Kernal, H., and Nass, C. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human-Computer Interaction*, 14(4):395–435.

[Nass, 2004] Nass, C. (2004). Etiquette equality: exhibitions and expectations of computer politeness. *Communications of the ACM*, 47(4):35–37.

[Nass et al., 2000] Nass, C., Isbister, K., and Lee, E.-J. (2000). Truth is beauty: researching embodied conversational agents. *Embodied Conversational Agents*, pages 374–402.

[Nass and Lee, 2001] Nass, C. and Lee, K. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, 7(3):171–181.

[Nass et al., 1999] Nass, C., Moon, Y., and Carney, P. (1999). Are respondents polite to computers? Social desirability and direct responses to computers. *Journal of Applied Social Psychology*, 29(5):1093–1110.

[Nass et al., 1997] Nass, C., Moon, Y., and Green, N. (1997). Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices. *Journal of Applied Social Psychology*, 27(10):864–876.

[Nass et al., 1994] Nass, C., Steuer, J., and Tauber, E. (1994). Computers are Social Actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'04)*, pages 72–78.

[Ndiaye et al., 2005] Ndiaye, A., Gebhard, P., Kipp, M., Klesen, M., Schneider, M., and Wahlster, W. (2005). Ambient Intelligence in Edutainment: Tangible Interaction with Life-Like Exhibit Guides. In *Proceedings of the International Conference on Intelligent Technologies for Interactive Entertainment*, LNCS 3814, pages 104–113. Berlin, Heidelberg, Germany: Springer.

[Nijholt et al., 2004] Nijholt, A., Rist, T., and Tuijnenbreijer, K. (2004). Lost in Ambient Intelligence? Panel Session. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'04)*, pages 1725–1726. New York, NY, USA: ACM.

[Norman, 1998] Norman, D. A. (1998). *Psychology of Everyday Things (reprinted as 'The Design of Everyday Things')*. New York, NY, USA: Bantam Books.

[Norman, 1999] Norman, D. A. (1999). Affordance, conventions, and design. *Interactions*, 6(3):38–43.

[North et al., 2004] North, A., MacKenzie, L., and Hargreaves, D. (2004). The effects of musical and voice fit on responses to advertisements. *Journal of Applied Social Psychology*, 34(8):1675–1708.

[Nowson, 2006] Nowson, S. (2006). *The Language of Weblogs: A study of genre and individual differences*. PhD thesis, University of Edinburgh, UK.

[Ohala, 1994] Ohala, J. (1994). *Sound Symbolism*, chapter The frequency codes underlies the sound symbolic use of voice pitch, pages 325–347. Cambridge, MA: Cambridge University Press.

[Omdahl, 1995] Omdahl, B. (1995). *Cognitive appraisal, emotion, and empathy*. Lawrence Erlbaum Associates.

[Overbeeke et al., 2004] Overbeeke, K., Djajadiningrat, T., Hummels, C., Wensveen, S., and Frens, J. (2004). *Funology: From Usability to Enjoyment*, chapter Let's make things engaging, pages 7–17. Amsterdam, The Netherlands: Kluwer.

[Oviatt, 1999] Oviatt, S. (1999). Mutual Disambiguation of Recognition Errors in a Multimodel Architecture. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'99)*, pages 576–583. New York, NY, USA: ACM.

[Pennebaker and King, 1999] Pennebaker, J. and King, L. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77:1296–1312.

[Persson et al., 2000] Persson, P., Laaksoiahti, J., and Lönnqvist, P. (2000). Anthropomorphism - a multi-layered phenomenon. In Dautenhahn, K., editor, *Proceedings of AAAI Fall symposium, Socially Intelligent Agents: The Human in the Loop*, pages 131– 135. AAAI Press.

[Persson et al., 2002] Persson, P., Laaksolahti, J., and Löonnqvist, P. (2002). *Socially Intelligent Agents*, chapter Understanding Social Intelligence, pages 21–28. Berlin, Heidelberg, Germany: Springer.

[Petty and Wegener, 1998] Petty, R. and Wegener, D. (1998). Attitude change: Multiple roles for persuasion variables. *The Handbook of Social Psychology*, 1:323–390.

[Piaget, 1933] Piaget, J. (1933). *A Handbook of Child Psychology*, chapter Children's philosophies, pages 505–516. Worcester, MA, USA: Clark University Press, 2nd edition.

[Plutchik, 1994] Plutchik, R. (1994). *The Psychology and Biology of Emotion*. New York, NY, USA: Harpercollins College Publishers.

[Povinelli and Bering, 2002] Povinelli, D. and Bering, J. (2002). The mentality of apes revisited. *Current Directions in Psychological Science*, 11:115–119.

[Poynor, 1995] Poynor, R. (1995). The hand that rocks the cradle. *ID - The International Design Magazine*, 42:60.

[Preece et al., 2002] Preece, J., Rogers, Y., and Sharp, H. (2002). *Interaction design - Beyond Human-Computer Interaction*. Hoboken, NJ, USA: Wiley.

[Pulkki, 1997] Pulkki, V. (June 1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466.

[Quinlan, 1993] Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann.

[Quraischy, 2009] Quraischy, H. (2009). Multimodale Interaktion mit virtuellen und realen Objekten am Beispiel eines Einkaufsassistenten für ländliche Ortschaften. Master's thesis, Saarland University, Saarbrücken, Germany.

[Reeps, 2004] Reeps, I. (2004). Joy of Use - eine neue Qualiät für interaktive Produkte. Master's thesis, Universität Konstanz, Germany.

[Reeves and Nass, 1996] Reeves, B. and Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge, MA: Cambridge University Press.

[Rocchesso et al., 2003] Rocchesso, D., Bresin, R., and Fernström, M. (2003). Sounding objects. *IEEE Multimedia*, 10(2):42–52.

[Rukzio et al., 2006] Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A., and Chin, J. (2006). An Experimental Comparison of Physical Mobile Interaction Techniques: Touching, Pointing and Scanning. In *Proceedings of the International Conference on Ubiquitous Computing (UbiComp 2006)*, pages 87–104. New York, NY, USA: ACM.

[Russell and Norvig, 2009] Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ, USA: Prentice hall.

[Ryokai et al., 2004] Ryokai, K., Marti, S., and Ishii, H. (2004). I/O Brush: Drawing with Everyday Objects as Ink. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'04)*, pages 303–310. New York, NY, USA: ACM.

[Saffer, 2009] Saffer, D. (2009). *Designing for Interaction: Creating Smart Applications and Clever Devices*. Berkeley, CA, USA: Peachpit Press.

[Sara et al., 2005] Sara, S.-L. L., Kiesler, S., Yee-man, I., and Chiu, L. C.-y. (2005). Human Mental Models of Humanoid Robots. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 2767–2772. Los Alamitos, CA, USA: IEEE Computer Society Press.

[Satyanarayanan, 2001] Satyanarayanan, M. (2001). Pervasive computing: vision and challenges. *IEEE Personal Communications*, 8(4):10 – 17.

[Scassellati, 2000] Scassellati, B. (2000). Discriminating Animate from Inanimate Visual Stimuli. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence*, pages 1405–1410. San Francisco, CA, USA: Morgan Kaufmann.

[Scherer, 1994] Scherer, K. (1994). *Emotions: Essays on emotion theory*, chapter Affect Bursts, pages 161–196. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

[Schmidt, 2000] Schmidt, A. (2000). Implicit human computer interaction through context. *Personal and Ubiquitous Computing*, 4(2):191–199.

[Schmidt, 2005] Schmidt, S. (2005). Persönlichkeitsaspekte in Stimmen von anthropomorphen Produkten. Diplomarbeit, Saarland University, Saarbrücken, Germany.

[Schmitz, 2003] Schmitz, M. (2003). Ein Framework für räumliches Audio in instrumentierten Umgebungen. Diplomarbeit, Saarland University, Saarbrücken, Germany.

[Schmitz, 2006] Schmitz, M. (2006). A Framework for Multi-User Support in Instrumented Public Spaces. In *Tagungsband der Mensch und Computer (MuC'06)*, pages 469–472. Munich, Germany: Oldenbourg Verlag.

[Schmitz, 2010] Schmitz, M. (2010). Designing Tangible Interaction With Anthropomorphic Smart Objects. In *Proceedings of the IADIS International Conference on Interfaces and Human Computer Interaction (IHCI 2010)*, pages 403–407. Lisbon, Portugal: IADIS press.

[Schmitz, 2011] Schmitz, M. (2011). Concepts for Anthropomorphism in Interactive Devices. In *Proceedings of the Fifth International Conference on Tangible, Embedded and Embodied Interaction (in print)*. New York, NY, USA: ACM.

[Schmitz et al., 2008a] Schmitz, M., Baus, J., and Dörr, R. (2008a). The Digital Sommelier: Interacting with Intelligent Products. In Floerkemeier, C., Langheinrich, M., Fleisch, E., Mattern, F., and Sarma, S. E., editors, *Proceedings of the First International Conference on The Internet of Things (IOT 2008)*, LNCS 4952, pages 247–262. Berlin, Heidelberg, Germany: Springer.

[Schmitz and Butz, 2006] Schmitz, M. and Butz, A. (2006). SAFIR: Low-Cost Spatial Audio for Instrumented Environments. In *Proceedings of the 2nd International Conference on Intelligent Environments*, pages 427–430. Amsterdam, The Netherlands: IOS Press.

[Schmitz et al., 2007a] Schmitz, M., Dörr, R., and Baus, J. (2007a). Interaction Patterns for Smart Products. In *Proceedings of the 1st Workshop on the Design of Smart Products (DoSP 2007)*, pages 13–21. Furtwangen, Germany: Research Center for Intelligent Media.

[Schmitz et al., 2008b] Schmitz, M., Endres, C., and Butz, A. (2008b). A Survey of Human-Computer Interaction Design in Science Fiction Movies. In *Proceedings of the International Conference on INtelligent TEchnologies for interactive enterTAINment (INTETAIN '08)*, pages 1–10. New York, NY, USA: ACM.

[Schmitz et al., 2007b] Schmitz, M., Krüger, A., and Schmidt, S. (2007b). Modelling Personality in Voices of Talking Products Through Prosodic Parameters. In *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI'07)*, pages 313–316. New York, NY, USA: ACM.

[Schmitz et al., 2009] Schmitz, M., Minina, E., and Quraischy, H. (2009). The Augmented Dorfladen – Interacting with Real and Virtual Products in Instrumented Shelves. In *Proceedings of 5th International Conference on Intelligent Environments*, pages 57–64. Amsterdam, The Netherlands: IOS Press.

[Schmitz and Quraischy, 2009] Schmitz, M. and Quraischy, H. (2009). Tangible Interaction with Real and Virtual Products: Designing a Shopping Assistant for Rural Communities. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction (TEI'09)*, pages 209–212. New York, NY, USA: ACM.

[Schmitz and Zimmer, 2007] Schmitz, M. and Zimmer, H. (2007). Big MaMUT is Watching You - Interaction-Tracking for Instrumented Environments. In *Proceedings of the 3rd International Conference on Intelligent Environments (IE'07)*, pages 451–454. Amsterdam, The Netherlands: IOS Press.

[Schneider, 2004] Schneider, M. (2004). Towards a Transparent Proactive User Interface for a Shopping Assistant. In *Proceedings of the Workshop on Multi-User and Ubiquitous User Interfaces (MU3I'04)*, pages 31–35.

[Schneider, 2010] Schneider, M. (2010). *Resource-Aware Plan Recognition in Instrumented Environments*. PhD thesis, Saarland University, Saarbrücken, Germany.

[Schneider and Kröner, 2008] Schneider, M. and Kröner, A. (2008). The Smart Pizza Packing: An Application of Object Memories. In *Proceedings of the International Conference on Intelligent Environments*, pages 1–8. Amsterdam, The Netherlands: IOS Press.

[Scholl and Tremoulet, 2000] Scholl, B. J. and Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8):299 – 309.

[Scholsberg, 1941] Scholsberg, H. (1941). A scale for the judgment of facial expressions. *Journal of Experimental Psychology*, 29(6):497–510.

[Schomaker et al., 1995] Schomaker, J., Nijtmans, A., Camurri, F. L., Morasso, P., Benoît, C., T. Guiard-Marigny, B., Le Goff, J., Robert-Ribes, A., Adjoudani, I., Defée, S., Münch, K., Hartung, and J. Blauert, A. (1995). A taxonomy of multimodal interaction in the human information processing system. Technical report, European Commission: CORDIS, ESPRIT Basic Research Action 8579.

[Schröder, 2001] Schröder, M. (2001). Emotional Speech Synthesis: A Review. In *Proceedings of the European Conference on Speech Communication and Technology*, pages 561–564. Grenoble, France: International Speech Communication Association.

[Schröder, 2003] Schröder, M. (2003). Experimental Study of Affect Bursts. *Speech Communications*, 40(1-2):99–116.

[Schröder and Trouvain, 2003] Schröder, M. and Trouvain, J. (2003). The german text-to-speech synthesis system mary: A tool for research, development and teaching. *International Journal of Speech Technology*, 6(4):365 – 377.

[Selye, 1970] Selye, H. (1970). Stress and aging. *Journal of the American Geriatrics Society*, 18(9):669.

[Sharlin et al., 2004] Sharlin, E., Watson, B., Kitamura, Y., Kishino, F., and Itoh, Y. (2004). On tangible user interfaces, humans and spatiality. *Personal Ubiquitous Computing*, 8(5):338–346.

[Shedroff, 1999] Shedroff, N. (1999). *Information Design*, chapter Information Interaction Design: A Unified Field Theory of Design, pages 267–292. Cambridge, MA: MIT Press.

[Sheridan et al., 2003] Sheridan, J. G., Short, B. W., Kortuem, G., Laerhoven, K. V., and Villar, N. (2003). Exploring Cube Affordance: Towards A Classification Of Non-Verbal Dynamics Of Physical Interfaces. In *Proceedings of IEE Eurowearable*, pages 113–118. Birmingham, UK: IEE Press.

[Shneiderman and Maes, 1997] Shneiderman, B. and Maes, P. (1997). Direct Manipulation vs. Interface Agents. *Interactions*, 4(6):42–61.

[Shoham, 1993] Shoham, Y. (1993). Agent-Oriented Programming. *Artificial Intelligence*, 60(1):51–92.

[Siebel, 2003] Siebel, N. T. (2003). *Design and Implementation of People Tracking Algorithms for Visual Surveillance Applications*. PhD thesis, The University of Reading, Reading, UK.

[Smith et al., 1975] Smith, B., Brown, B., W.Strong, and Rencher, A. (1975). Effects of speech rate on personality perception. *Language and Speech*, 18(2):145–152.

[Stahl, 2009] Stahl, C. (2009). *Spatial Modeling of Activity and User Assistance in Instrumented Environments*. PhD thesis, Saarland University, Saarbrücken, Germany.

[Stahl et al., 2005] Stahl, C., Baus, J., Brandherm, B., Schmitz, M., and Schwartz, T. (2005). Navigational- and Shopping Assistance on the Basis of User Interactions in Intelligent Environments. In *Proceedings of the IEE International Workshop on Intelligent Environments (IE 2005)*, University of Essex, Colchester, UK.

[Stanton et al., 2001] Stanton, D., Bayon, V., Neale, H., Ghali, A., Benford, S., Cobb, S., Ingram, R., O'Malley, C., Wilson, J., and Pridmore, T. (2001). Classroom Collaboration in the Design of Tangible Interfaces for Storytelling. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'01)*, pages 482–489. New York, NY, USA: ACM.

[Stauffer and Grimson, 1999] Stauffer, C. and Grimson, W. E. L. (1999). Adaptive Background Mixture Models for Real-Time Tracking. In *Proceedings of the IEEE Computer Science Conference on Computer Vision and Pattern Recognition*, pages 246–252. Los Alamitos, CA, USA: IEEE Computer Society Press.

[Stephens, 1997] Stephens, R. (1997). A survey of stream processing. *Acta Informatica*, 34(7):491–541.

[Stipek, 1996] Stipek, D. (1996). *Handbook of Educational Psychology*, chapter Motivation and Instruction, pages 85–113. London, UK: Prentice Hall.

[Streeter et al., 1977] Streeter, L., Krauss, R., Geller, V., Olson, C., and Apple, W. (1977). Pitch changes during attempted deception. *Journal of Personality and Social Psychology*, 35(5):345–350.

[Taffin, 2004] Taffin, D. P. (2004). The Tangible User Interface ”Flip'n Twist”. Bachelor's thesis, Saarland University, Saarbrücken, Germany.

[Tan and Fasting, 1996] Tan, E. and Fasting, B. (1996). *Emotion and the structure of narrative film: Film as an emotion machine*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

[Togler et al., 2009] Togler, J., Hemmert, F., and Wettach, R. (2009). Living Interfaces: The Thrifty Faucet. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pages 43–44. New York, NY, USA: ACM.

[Ullmer and Ishii, 2001] Ullmer, B. and Ishii, H. (2001). *Human-Computer Interaction in the New Millennium*, chapter Emerging Frameworks for Tangible User Interfaces, pages 579–601. Essex, UK: Addison Wesley.

[Ullmer et al., 1998] Ullmer, B., Ishii, H., and Glas, D. (1998). mediaBlocks: Physical Containers, Transports, and Controls for Online Media. In *Proceedings of the 25th Conference on Computer Graphics and Interactive Techniques*, pages 379–386. New York, NY, USA: ACM.

[Underkoffler and Ishii, 1999] Underkoffler, J. and Ishii, H. (1999). Urp: A Luminous-Tangible Workbench for Urban Planning and Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'99)*, pages 386–393. New York, NY, USA: ACM.

[van Dam, 1997] van Dam, A. (1997). Post-WIMP User Interfaces. *Communications of the ACM*, 40(2):63–67.

[Wada and Shibata, 2007] Wada, K. and Shibata, T. (2007). Living with seal robots - its sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics*, 23(5):972–980.

[Wahlster, 2003] Wahlster, W. (2003). Towards Symmetric Multimodality: Fusion and Fission of Speech, Gesture, and Facial Expression. *Advances in Artificial Intelligence*, LNCS 2821:1–18.

[Wahlster, 2006] Wahlster, W. (2006). *SmartKom: Foundations of Multimodal Dialogue Systems (Cognitive Technologies)*. Berlin, Heidelberg, Germany: Springer.

[Wahlster et al., 2008] Wahlster, W., Kröner, A., Schneider, M., and Baus, J. (2008). Sharing Memories of Smart Products and their Consumers in Instrumented Environments (Austausch digitaler Gedächtnisse von Produkten und Konsumenten in instrumentierten Umgebungen). *it - Information Technology*, 50(1):45–50.

[Wasinger, 2006] Wasinger, R. (2006). *Multimodal Interaction with Mobile Devices: Fusing a Broad Spectrum of Modality Combinations*, volume 305 of *DISKI*. Berlin, Germany: Akademische Verlagsgesellschaft Aka.

[Wasinger et al., 2005] Wasinger, R., Krüger, A., and Jacobs, O. (2005). Integrating Intra and Extra Gestures into a Mobile and Multimodal Shopping Assistant. In *Proceedings of the 3rd International Conference on Pervasive Computing*, pages 297–314. New York, NY, USA: ACM.

[Wasinger and Wahlster, 2005] Wasinger, R. and Wahlster, W. (2005). The Anthropomorphized Product Shelf: Symmetric Multimodal Human-Environment Interaction. In *True Visions: Tales on the Realization of Ambient Intelligence*, pages 291–306. Berlin, Heidelberg, Germany: Springer.

[Weiser, 1991] Weiser, M. (1991). The computer for the 21st century. *Scientific American*, 265(3):94–104.

[Wellner et al., 1993a] Wellner, P., Mackay, W., and Gold, R. (1993a). Back to the real world. *Communications of the ACM*, 36(7):24–27.

[Wellner et al., 1993b] Wellner, P., Mackay, W. E., and Gold, R. (1993b). Computer-Augmented Environments: Back to the Real World - Introduction to the Special Issue. *Communications of the ACM*, 36(7):24–26.

[White, 1995] White, P. (1995). *The understanding of causation and the production of action: From infancy to adulthood*. London, UK: Psychology Press.

[Winograd, 1997] Winograd, T. (1997). *From Computing Machinery to Interaction Design*, volume Beyond Calculation: The Next Fifty Years of Computing, pages 149–162. Berlin, Heidelberg, Germany: Springer.

[Witten and Frank, 2005] Witten, I. H. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, CA, USA: Morgan Kaufmann.

[Wooldridge and Jennings, 1995] Wooldridge, M. and Jennings, N. R. (1995). Intelligent Agents: Theory and Practice. *Knowledge Engineering Review*, 10(2):115–152.

[Zimmer, 2006] Zimmer, H. (2006). *Multi-User Tracking in Instrumented Rooms*. Bachelor's thesis, Saarland University, Saarbrücken, Germany.

[Zuckerman and Miyake, 1993] Zuckerman, M. and Miyake, K. (1993). The attractive voice: What makes it so? *Journal of Nonverbal Behavior (Historical Archive)*, 17(2):119 – 135.

[Zuehlke, 2010] Zuehlke, D. (2010). SmartFactory-Towards a Factory-of-Things. *Annual Reviews in Control*, 34(1):129 – 138.