A Wizard-of-Oz Setting for Multimodal Interaction

An Approach to User-Based Elicitation of Design Patterns*

Andreas Ratzka

Universität Regensburg
Institut für Medien-, Informations- und Kulturwissenschaft
93040 Regensburg
Andreas.Ratzka@sprachlit.uni-regensburg.de

Zusammenfassung
Dieser Beitrag beschreibt ein Wizard-of-Oz-Setting für multimodale Interaktion. Dieser Versuchsaufbau ist ein wichtiger Arbeitsbaustein für die Entwicklung einer Entwurfsmethodik für multimodale Interaktion auf Basis von Wiederverwendung und Nutzerpartizipation. Wiederverwendung von Entwürfen geschieht hierbei mit Hilfe von *Design Pattern Languages*, einer Methode zur Verwaltung von Designinformationen für Entwickler und Usability-Experten. Nutzerpartizipation erfolgt über Wizard-of-Oz-Tests, bei denen das Systemverhalten in frühen Entwicklungsphasen durch menschliche Agenten simuliert wird. Als Testszenario dient die Interaktion mit einem multimodalen E-Mail-System für mobile Geräte und für Desktoprechner. Dabei sollen Hypothesen über multimodale Interaktion erkundet und Design-Pattern-Kandidaten identifiziert werden.

Abstract
This paper describes a Wizard of Oz setting for multimodal interaction. The design of this experimental setup is one working package in our research on a design methodology for multimodal interaction based on design reuse and user participation. Design reuse can be accomplished by design pattern languages, which are a powerful means of information management for both system developers and usability experts. User involvement is performed via Wizard-of-Oz tests, which simulate system behaviour by means of human agents. The test scenario is managing e-mail with multimodal mobile and desktop devices. Our goal is to explore hypotheses on mul-

---

* Published in: Osswald, Achim; Stempfhuber, Maximilian; Wolff, Christian (eds.) (2007). Open Innovation. Proc. 10th International Symposium for Information Science. Constance: UVK, 159-170.

timodal interaction and thus identify pattern candidates for the design of multimodal interaction.

1 Introduction

This paper describes the research design for Wizard of Oz simulations of multimodal interactive systems and is situated in the context of methodological research on multimodal interaction design. After a thorough review of literature on model-based user interface design (such as Bürgy 2002, Calvary et al. 2003, da Silva 2001, Souchon et al. 2002, Trætteberg 2002, Wilson et al. 1993) following design aspects for multimodal interaction have been presented in previous work (Ratzka 2006, Ratzka & Wolff 2006):

- Cross-application-aspects
- Workflow and task
- Contextual aspects, i.e. user, situation, environment, device
- Adaptation concept
- Application-specific aspects
- Interaction
- Presentation
- Software Architecture

Our methodology is based on design reuse and user participation. Design reuse can be accomplished with so called *design patterns* and *pattern languages*. User participation and early user involvement is done with Wizard of Oz tests.

The term *design pattern* was firstly used by Alexander et al. (1977) in the domain of architecture and later on introduced into the fields of software engineering (Gamma et al. 1995, Trowbridge & Cunningham 2001) and user interface design (Borchers 2001, van Welie & van der Weer 2003). Design patterns describe solutions to well known design problems in a structured and retrievable way. They relate together a striking name, context and problem descriptions, the description of a rather abstract solution, concrete examples, as well as strengths and pitfalls of this solution.

Patterns constitute an information management tool for system and interaction design, especially when they are grouped together as interlinked pattern language, which can be easily explored by the system designer. Furthermore, by means of pattern languages, common design problems and their solutions can be called by their name and can be discussed more easily by both software and usability experts.

Multimodality has proven to be a successful solution, especially for map based tasks and image editing (Oviatt 1996, Cohen et al. 1997, 2000, Ren et al. 2000, Raisamo & Räihä 2000, Gorniak & Roy 2003, Hiyoshi & Shimazu 1994, Milota 2004). Research on multimodal interaction with automotive applications (Neuss 2001, Salmen 2002, Seiffert 2002, Niedermaier 2003) and personal assistants (Bers et al. 1998, Comeford et al. 2001, Miyazaki 2002) indicate the plausibility of multimodality in these contexts as well.

Some of the high level research results gained in these domains, such as the often cited buzzwords *mutual disambiguation* and *redundant information display*, can be formulated as reusable solutions to recurring design problems and thus be described in a pattern format. In previous work, we have identified a still incomplete collection of patterns and pattern candidates for multimodal interaction from literature review (cf. Ratzka & Wolff 2006). The pattern *human action source distribution* will serve as an illustrative example:

1.1 Human Action Source Distribution

Context The input of different data types (common in image editing systems, such as selecting a tool from a palette and drawing a figure) sometimes requires a repetitive repositioning of the user's hands between mouse and keyboard or a repetitive repositioning of the mouse cursor.

Problem Repetitive repositioning is non-productive and slows down interaction.

Solution Partition alternating subtasks into different action channels. Make use of two-handed (Raisamo & Räihä 2000) or speech-enhanced multimodal interaction (Hiyoshi & Shimazu 1994, Milota 2004).

Design patterns relate together problems and *well proven* solutions. In contrast to WIMP[1]-based interaction and web-based applications, multimodality has not yet reached a high degree of dissemination in consumer electronics. Only few design questions can be countered with well proven solutions from research. Other research results hold for specialized contexts only and cannot be generalised. Pattern candidates won from them are proven solutions only in that context. This imposes that user participation is needed to verify whether a certain pattern really fits into the target context, whether a pattern candidate can be generalised to a valid pattern. Furthermore multimodal interaction invites to new design ideas which seem plausible but are not proven empirically. That's why our literature review has to be complemented by user involvement, performed by means of Wizard of Oz tests.

---

[1] WIMP: **W**indows, **I**cons, **M**enus, **P**ointing device.

2 Experimental Setting

Wizard of Oz simulations are a well known approach of informal prototyping in HCI research, especially in spoken interaction and multimodal interaction (cf. Balbo et al. 1993, Salber & Coutaz 1993, Womser-Hacker 1993, Hitzenberger & Womser-Hacker 1995, Cenek et al. 2005, Klemmer et al. 2000). They tend to counter a common problem of human computer interaction researchers when exploring new interface styles: At early stages of interface design, necessary components for interpreting user input are not yet available. The requirements needed to implement them are still unknown, but can be elicited by early user participation accomplished by means of simulations.

Thus Wizard of Oz settings allow user participation at the very first stages of interface design. Alternative approaches of informal prototyping comprise paper mockups and storyboarding, which are both exploited in multimodal user interface design (Chandler et al. 2002, Sinha & Landay 2001, Sinha & Landay 2003). In contrast to those techniques, Wizard of Oz tests provide to test users a more realistic look and feel so that design alternatives can be discussed more deeply.

To perform Wizard tests, you need at least two connected computers, one for the user interface, for the test person to perform interactive tasks, and another one for the Wizard, the person who interprets user input and simulates system behaviour. In multimodal scenarios, multiple Wizards may be required, e. g. one for input and one for output or one for each input modality.

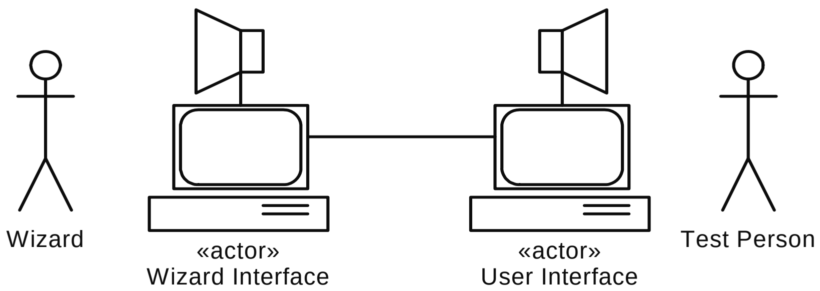

Fig. 1: Wizard of Oz System Setup

This work focuses on the multimodal enhancement of WIMP-based desktop and mobile interaction with spoken elements. Whereas the WIMP-based behaviour is implemented directly, spoken input and multimodal fusion have to be simulated. The wizard interprets user input and may change the data model of the application accordingly, so that the view of the user interface is updated. He can trigger preformulated spoken output, unless an automatic speech feedback strategy is activated. This requires an efficient wizard interface, which can be controlled in real time via shortcut keys.

The platform simulates a mobile multimodal digital assistant (mobile setting) as well as a multimodal desktop e-mail client (desktop setting). The test person has to perform the task provided by the test scenario, i.e. to organise her incoming mail, organize and delegate meetings etc.

This scenario has been chosen because e-mail is one of the most successful and widespread computer applications (Ducheneaut & Watts 2005, Whittaker et al. 2005). Furthermore, other research (Bers et al. 1998, Comeford et al. 2001, Miyazaki 2002) indicates the plausibility of multimodal interaction for this application domain. In three consecutive surveys (cognitive walkthrough with usability experts, qualitative user study, quantitative user study) hypotheses on multimodal interaction will be examined. In these studies, well determined test scenarios (in the style of *mail baskets*, which are popular exercises in assessment centres) will be combined with free exploration of the user interface by the test person.

One central question to be answered via this method concerns the feedback strategy, especially the interplay and balance between graphical and acoustic system output. Below the two pattern candidates *system initiated modality hint* and *speech-enhanced display* are shortly outlined as examples. The former attempts to counter the conflict between data appropriate modality allocation (such as graphical display of long lists) and interaction history appropriate modality allocation (answering to speech commands via speech replies). The latter one picks up the potential of multimodal interaction to amplify virtually the display of a small device via speech output, so that cluttered screens and navigation complexity can be minimised.

2.1 System Initiated Modality Hint

Context Multimodality provides the possibility to select modalities flexibly according to task and data properties, user preferences, and context factors.

Problem Sometimes data properties and context factors are in conflict, that means, they encourage the selection of different modalities. Both system and user may lack information for selecting the most feasible interaction style, so that neither pure system initiative nor pure user initiative seem appropriate for modality selection.

Solution Point out alternative interaction styles to the user via spoken hints. If the user requests information via speech commands (“get messages from John Long”) and the search result is more easily displayed graphically (as a list of messages) a spoken modality hint (such as "look at the display or say 'read out'") would complement the visual output but preserve interaction flexibility.

2.2 Speech-enhanced Display

Context	Some interactive systems support a lot of different tasks. During daily work it is necessary to switch between tasks to get the information needed to perform the main task. When a user wants to answer an e-mail he might need other data which can be found in previous correspondence.
Problem	Small displays make such "multi-tasking" very difficult. Different task windows cannot be displayed all at once, whereas switching between several task screens imposes high navigational efforts, such that users might loose orientation.
Solution	If the display is too small for simultaneous presentation of several task windows, consider partitioning different task displays to different output modalities. Users might listen to new mail or to e-mails of a specific thread while composing a message on the same topic and preserving the visual display.

Both pattern candidates seem plausible but still lack empirical evidence which can be obtained only by user participation. In contrast to full functioning prototypes, Wizard of Oz simulations can react more flexibly to incrementally revealed interaction requirements which are unknown in the very first phases of development.

3 Application Architecture

From the perspective of software architecture human computer interaction can be viewed as the creation, selection, manipulation, and destruction of interaction objects. Multimodal interaction imposes the need of specialized interaction objects for covering modality specific interaction possibilities. In order to maintain consistency and to enable synergistic and alternating modality usage, these specialized interaction objects have to be inter-connected through standardized interaction protocols.

That's why the interaction objects of the (simulated) user interfaces are organized in a distributed multi-agent architecture[2] making use of observer, adapter, proxy, and other design patterns. This architecture is inspired by the MVC-extensions of Visual Works Smalltalk (Krasner & Pope 1988, Lewis 1995 p. 103 ff.) on the one hand and PAC-Amadeus (Coutaz 1987, Coutaz 1994, Nigay & Coutaz 1995) on the other hand.

In MVC- and PAC-based architectures observer-objects (here: views) are wrapped around each functional core object (here: model) to get notified about relevant data changes, i.e. to maintain consistency between the functional core and the user interfaces.

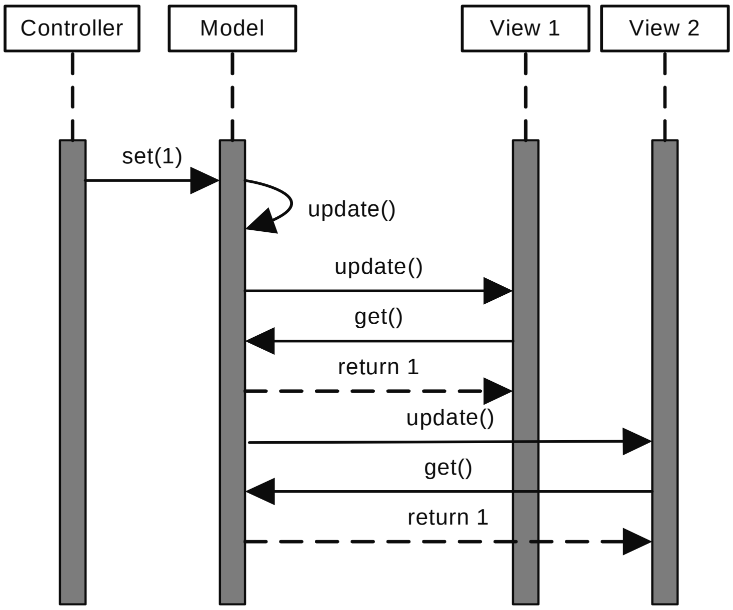

Fig. 2: The Observer Pattern in the MVC Paradigm

Adapters simplify the complex interface of some functional core objects to fit to the MVC protocol.

The proxy-pattern is used in order to introduce one or more degrees of indirection between objects belonging to the functional core of the application and those belonging to the user interface component.

Indirection is required when several interaction objects are combined together, e. g. one for selecting an interaction object and another one for modifying the current selection. This combination of the observer and proxy patterns is also called *subject channel* (view Lewis 1995 p. 103 ff.).

2 In our case we mean reactive agents and not intelligent agents (cf. Coutaz 1994).

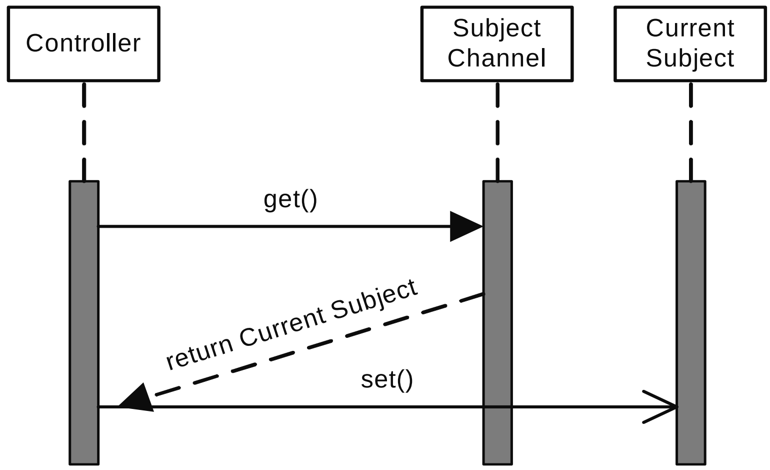

Fig. 3: The Proxy Pattern in extended MVC Architectures with Subject Channels

Further indirection is required to make interaction objects remote-controllable which is needed in a distributed Wizard of Oz scenario but also in real working systems, where the speech recognition component and the user interface are running on separate computers.

4 Technological Base

The Wizard of Oz platform is realized as a distributed system consisting of a pocket PC and a desktop PC (mobile setting) or of two desktop PCs (desktop setting) respectively. The communication between the distributed components is realized via socket connections through USB, Bluetooth or Ethernet.

The platform is implemented using *ewe* (cf. Brereton 2005), a Java-based toolkit providing libraries for programming highly portable java applications. The application classes can be run with an ordinary J2SE runtime environment or, and that is necessary for mobile applications, using the Ewe virtual machine.

Applications to be run with the Ewe virtual machine are restricted to the use of the Ewe API (including a subset of java.lang classes), and third party libraries, e. g. your own code.

Applications run with a J2SE runtime environment can combine both standard Java and Ewe libraries. The resulting applications are less portable, but this approach provides more flexibility for implementing the desktop components, i.e. the Wizard interface and the user interface for the desktop setting using JFC (cf. Sun Microsystems 2006), which are easier to handle and more thoroughly documented than the Ewe GUI library.

5 Conclusion and Future Work

This paper describes a Wizard of Oz setting in the context of methodological research on multimodal interaction for mobile and desktop devices. Early user involvement performed via wizard tests complements the reuse-based approach of using design patterns. This is important because multimodal interaction is still innovative and therefore lacks a large inventory of proven design solutions. User involvement is needed in design phases as early as possible, which encourages the application of Wizard of Oz simulations.

This research intends to examine design aspects concerning task, context of use, adaptation, interaction and presentation in more detail by means of Wizard of Oz simulations.

In the short run we will perform qualitative user tests in order to validate the general plausibility of design decisions motivated by our pattern candidates. These user tests will be complemented by open interviews and focus group discussions in order to elicit further pattern candidates.

In the long term evaluation tools, eye-tracking as additional interaction modality, as well as a speech engine to allow more scalable user interface simulations have to be integrated in order to allow the collection of reliable quantitative data.

As a by-product, architectural considerations relevant for the design of multimodal user interfaces, which become apparent when implementing Wizard of Oz prototypes, will be collected and further systematised.

6 References

Alexander, C.; Ishikawa, S.; Silverstein, M.; Jacobson, M.; Fiksdahl-King, I.; Angel, S. (1977). *A Pattern Language.* Oxford University Press.

Balbo, S.; Coutaz, J.; Salber, D. (1993). Towards automatic evaluation of multimodal user interfaces. In: *IUI '93: Proceedings of the 1st international conference on Intelligent user interfaces.* New York, NY, USA: ACM Press, 201-208.

Bers, J.; Miller, S.; Makhoul, J. (1998). Designing conversational interfaces with multimodal interaction. In: *DARPA Workshop on Broadcast News Understanding Systems.* DARPA, 319-321.

Borchers, J. O. (2001). A Pattern Approach to Interaction Design. *AI & Society Journal of Human-Centred Systems and Machine Intelligence 15*(4), Springer, 359-376.

Brereton, M. (2005). *Welcome to ewe. The easiest way to program your mobile PC.* Website: Ewesoft, December 20th 2005. Lastly retrieved: April 04, 2007 from: http://www.ewesoft.com/.

Bürgy, C. (2002). An Interaction Constraints Model for Mobile and Wearable Computer-Aided Engineering Systems in Industrial Applications. Doctoral Thesis, University of Pittsburgh, Pennsylvania, USA .

Calvary, G.; Coutaz, J.; Thevenin, D.; Limbourg, Q.; Bouillon, L.; Vanderdonckt, J. (2003). A unifying reference framework for multi-target user interfaces. *Interacting with Computers 15*(3), 289-308.

Cenek, P.; Melichar, M.; Rajman, M. (2005). A Framework for Rapid Multimodal Application Design. *Proc. of TSD'05.* Karlovy Vary, Czech Republic, Springer, 393-403.

Chandler, C. D.; Lo, G.; Sinha, A. K. (2002). Multimodal theater: extending low fidelity paper prototyping to multimodal applications. In: *CHI '02: CHI'02 extended abstracts on Human factors in computing systems.* New York, NY, USA: ACM Press.

Cohen, P. R.; Johnston, M.; McGee, D.; Oviatt, S.; Pittman, J.; Smith, I.; Chen, L.; Clow, J. (1997). QuickSet: multimodal interaction for distributed applications. In: *MULTIMEDIA '97: Proceedings of the fifth ACM international conference on Multimedia.* New York, NY, USA: ACM Press, 31-40.

Cohen, P.; McGee, D.; Clow, J. (2000). The efficiency of multimodal interaction for a map-based task. In: *Proceedings of the sixth conference on Applied natural language processing.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 331-338.

Comerford, L.; Frank, D.; Gopalakrishnan, P.; Gopinath, R.; Sedivy, J. (2001). The IBM Personal Speech Assistant. In: *Proc. of IEEE ICASSP'01.* DARPA, 319-321.

Coutaz, J. (1987). PAC, an object-oriented model for dialog design. In: *Proceedings of IFIP INTERACT'87: Human-Computer Interaction*, chapter 2: Design and Evaluation Methods: 2.5 Dialogue Design and Evaluation, 431-436.

Coutaz, J. (1994). *Software Architecture Modeling For User Interfaces.* In: J. J. Marciniak, (ed.): *Encyclopedia of Software Engineering.* Wiley, Chichester, 38-49.

Ducheneaut, N.; Watts, L. A. (2005). In Search of Coherence: A Review of E-Mail Research. *Human Computer Interaction 20*, 1-9.

Gamma, E.; Helm, R.; Johnson, R.; Vlissides, J. (1995). *Design Patterns: Elements of Reusable Object-Oriented Software.* Addison-Wesley.

Gorniak, P.; Roy, D. (2003). Augmenting user interfaces with adaptive speech commands. In: *ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces.* New York, NY, USA: ACM Press, 176-179.

Grasso, M. A.; Ebert, D. S.; and Finin, T. W. (1998). The integrality of speech in multimodal interfaces. *ACM Transactions on Compuer-Human Interaction 5*(4), 303-325.

Hitzenberger, L.; Womser-Hacker, Ch. (1995). Experimentelle Untersuchungen zu multimodalen natürlichsprachlichen Dialogen in der Mensch-Computer-Interaktion. *SDV – Sprache und Datenverarbeitung. 19*(1), 51-61.

Hiyoshi, M.; Shimazu, H. (1994). Drawing pictures with natural language and direct manipulation. In: *Proceedings of the 15th conference on Computational linguistics.* Morristown, NJ, USA: Association for Computational Linguistics, 722-726.

Klemmer, S. R.; Sinha, A. K.; Chen, J.; Landay, J. A.; Aboobaker, N.; Wang, A. (2000). Suede: a Wizard of Oz prototyping tool for speech user interfaces. In: *UIST '00: Proceedings of the 13th annual ACM symposium on User interface software and technology.* New York, NY, USA: ACM Press 1-10

Krasner, G. E. and Pope, S. T. (1988). A cookbook for using the model-view-controller user interface paradigm in Smalltalk-80. *Journal of Object Oriented Programming* 1(3), 26-49.

Lewis, S. (1995). *The Art and Science of Smalltalk.* Prentice Hall: London et al.

Milota, A. D. (2004). Modality fusion for graphic design applications. In: *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces.* New York, NY, USA: ACM Press, 167-174.

Miyazaki, J. (2002). Discussion Board System with modality variation: From Multimodality to User Freedom. Tampere University, Diploma Thesis.

Neuss, R. (2001). *Usability Engineering als Ansatz zum Multimodalen Mensch-Maschine-Dialog.* Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, Doctoral Thesis.

Niedermaier, F. B. (2003). *Entwicklung und Bewertung eines Rapid-Prototyping Ansatzes zur multimodalen Mensch-Maschine-Interaktion im Kraftfahrzeug.* Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München, Doctoral Thesis.

Nigay, L.; Coutaz, J. (1995). A generic platform for addressing the multimodal challenge. In: *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems 1: Multimodal Interfaces*. 98-105.

Oviatt, S. (1996). Multimodal Interfaces for dynamic interactive maps. In: *Human Factors in Computing Systems, CHI '96 Conference Proceedings.* ACM Press, 95-102

Raisamo, R.; Räihä, K.-J. (2000). Design and evaluation of the alignment stick. In: *Interacting with Computers 12*(5), 483-506.

Ratzka, A.; Wolff, Ch. (2006). A Patern-based Methodology for Multimodal Interaction Design In: *Proc. TSD 2006.* Springer, 677-686.

Ratzka, A. (2006). Combining Modality Theory and Context Models. In: *Proc. PIT 2006.* Springer, 141-151.

Ren, X.; Zhang, G.; Dai, G. (2000). An Experimental Study of Input Modes for Multimodal Human-Computer Interaction. In: *ICMI '00: Proceedings of the Third International Conference on Advances in Multimodal Interfaces.* London, UK: Springer-Verlag, 49-56.

Salber, D.; Coutaz, J. (1993). A Wizard of Oz platform for the study of multimodal systems. In: *CHI '93: INTERACT '93 and CHI '93 conference companion on Human factors in computing systems.* New York, NY, USA: ACM Press, 95-96.

Salmen, A. (2002). *Multimodale Menüausgabe im Fahrzeug.* Lehrstuhl für Informationswissenschaft, Philosophische Fakultät IV, Universität Regensburg, Doctoral Thesis.

Seifert, K. (2002). Evaluation multimodaler Computer-Systeme in frühen Entwicklungsphasen, Fakultät V – Verkehrs- und Maschinensysteme. Technische Universität Berlin, Doctoral Thesis.

da Silva, P. P. (2001). User Interface Declarative Models and Development Environments: A Survey. In: P. Palanque and F. Patern (Eds.): *Proc. DSV-IS'00.* LNCS 1946, Springer, 207-226.

Sinha, A. K.; Landay, J. A. (2001). Visually prototyping perceptual user interfaces through multimodal storyboarding. In: *PUI '01: Proceedings of the 2001 workshop on Perceptive user interfaces.* New York, NY, USA: ACM Press, 1-4.

Sinha, A. K.; Landay, J. A. (2003). Capturing user tests in a multimodal, multidevice informal prototyping tool. In: *ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces.* New York, NY, USA: ACM Press.

Souchon, N.; Limbourg, Q.; and Vanderdonckt, J. (2002). Task Modelling in Multiple Contexts of Use. In: P. Forbrig et al.(Eds.): *Proc. DSV-IS'02.* LNCS 2545, Springer, 59-73.

Sun Microsystems (2006). *The Java™ Tutorials. Trail: Creating a GUI with JFC/SWING.* Lastly retrieved: April 04, 2007 from: http://java.sun.com/docs/books/tutorial/uiswing/.

Trætteberg H. (2002). *Model-based User Interface Design.* Doctoral Thesis, Department of Computer and Information Science, Norwegian University of Science and Technology.

Trowbridge, D.; Cunningham, W.; Evans, M.; Brader, L.; Slater, P. (2004). *Describing the Enterprise Architectural Space.* Microsoft Corporation, June 2004.

Van Welie, M.; van der Weer, G. C. (2003). Pattern Languages in Interaction Design: Structure and Organization. In: *Proc. Interact'03.*

Whittaker, S.; Bellotti, V.; Moody, P. (2005). Introduction to This Special Issue on Revisiting and Reinventing E-Mail. *Human Computer Interaction 20*, 1-9.

Wilson, S.; Johnson P.; Kelly, C.; Cunningham, J.; and Markopoulos P. (1993). Beyond Hacking: a Model Based Approach to User Interface Design. *In: Proc. HCI'93.*

Wolff, Ch. (2005). Media Design Patterns. In: Eibl, M.; Womser-Hacker, Ch. & Wolff, Ch. (Eds.). *Designing Information Systems.* Constance: UVK, 209-217.

Womser-Hacker, Ch. (1993). Statistical Experiments on Computer Talk. In: *First International Conference on Quantitative Linguistics.* Dordrecht et al.: Kluwer, 251-263.