Project ref. no.	FP6-507609
Project acronym	SIMILAR
Deliverable status	R
Contractual date of delivery	28 February 2007
Actual date of delivery	18 October 2007
Deliverable number	D100
Deliverable title	Iteration on Multimodal Usability Book
Nature	Report
Status & version	Final
Number of pages	8
WP contributing to the deliverable	SIG7
WP / Task responsible	NISLab
Editor	N/A
Author(s)	Niels Ole Bernsen and Laila Dybkjær
EC Project Officer	Mats Ljungqvist
Keywords	Book draft comments, book draft revision.
Abstract (for dissemination)	This deliverable documents work done to gather input for revising the first draft of the SIMILAR-instigated book on Multimodal Usability as well as the ensuing plans for revising the book.

SIMILAR project (IST - FP6-507609)

SIMILAR

Deliverable D100

Iteration of Multimodal Usability Book

18 October 2007

Authors

Niels Ole Bernsen[1] *and Laila Dybkjær*[1]

1: NISLab, Odense, Denmark


similar
network of excellence

Table of Contents

1 This Report

This report documents work done to gather input for revising the first draft of a SIMILAR-instigated book on Multimodal Usability as well as the ensuing plans for revising the book.

In autumn 2005, it was agreed within SIMILAR that the authors of the present report would write a draft manuscript on how to work with users in testing the many multimodal research prototype systems that were emerging from work carried out in the Network. The draft manuscript would then be circulated to all interested parties in SIMILAR who would either provide their comments upon reading the book, use it for guidance when testing the usability of their own systems and report on the results, or both.

The draft manuscript was completed in summer of 2006 and circulated within SIMILAR. This resulted in numerous valuable comments on the manuscript from SIMILAR partners. The comments are summarised in Section 3. To make sure that the book was also applied to some of the research prototype multimodal systems developed in SIMILAR, the authors undertook to test the usability of several systems themselves. This was done in June 2007 and the usability test results are reported in SIMILAR Deliverable D98 *Report on Iterative Testing of Multimodal Usability and Evaluation Guide*. However, D98 does not report on the lessons learned for the purpose of improving the draft manuscript, which will be done in the present report (Section 5).

Meanwhile, the draft manuscript had been submitted to Springer's Book Series on Human-Computer Interaction (HCI). Two substantial, collated book reviews written by experts on usability were received by the authors in summer of 2007, providing additional and important feedback on the draft manuscript. These reviews are summarised in Section 4. In response to the mentioned feedback from multiple sources on the draft manuscript, the authors wrote a manuscript revision plan which has been accepted by Springer. The plan is presented in Section 6.

Below, we start with a brief overview of the structure of the draft manuscript (Section 2).

2 Manuscript Structure and Contents

Very briefly, in order to make the discussion in the following sections intelligible to readers, the draft manuscript has the following tri-partite structure. We *first* introduce, in Chapters 3 and 4, two theoretical approaches, one about the factors relevant to system usability and its evaluation, the second about modalities (i.e., modality theory). These two theories correspond to the two elements in the title of the manuscript, i.e. *Multimodal Usability*. Secondly, in Chapter 5, we apply the theoretical apparatus to development for usability. Thirdly, in Chapters 6 through 10, we present and discuss methods, techniques, and knowledge sources for usability development and evaluation and handling of the collected data. The first two chapters are the introduction (Chapter 1) and a brief introduction to software engineering (Chapter 2).

3 Comments Received from Colleagues in SIMILAR

The manuscript comments received from eight SIMILAR colleagues are briefly summarised and discussed below under three headings: positive comments which do not suggest any action taken (Section 3.1); suggested modifications (Section 3.2); and suggested additions (Section 3.3).

3.1 Positive

1. Innovative on multimodality.
2. Generally clear, easy to understand.
3. Helpful for beginners and experts.
4. Comprehensive.
5. Lots of practical advice.
6. The humour is OK.
7. Good in structure and content.
8. Will enrich the literature around on that topic very much.

3.2 Suggested Modifications

1. **Concepts**: clarify the basic usability/parameter/property/value hierarchy.
2. **Software engineering**: more on usability in the life-cycle, especially on the roles of usability development vs. usability evaluation.
3. **Theory and practice**: link modality theory to the practical sections.
4. **Theory**: expand on input modality theory.
5. **Usability and functionality**: clarify even more.
6. **Development:** add more concrete hints on how to.
7. **Readers' guide**: relate contents to needs of different groups of readers.
8. **Presentation**: more succinct in places.
9. **Layout**: less dense text layout.
10. **Index:** add one.

These points are all well-taken. **(1)** requests further clarification of partially new theory on usability factors introduced in the manuscript. This should also take care of Point **(5)**. **(3)** points to the fact that, in the first version of the manuscript, our main emphasis was to get some practical advice out to developers in SIMILAR rather than to make this strongly required linking which, if successful, will provide half of the "backbone" of the book, the other half being addressed in Point (1). **(4)** is much sought-after by developers.

Points **(2)** and **(6)** address difficulties in the draft manuscript to do with whether, how, and to which extent to discuss software engineering and how to present development for usability in a maximally practically useful manner, respectively. Many readers do not need to be told about software engineering at all, but some do, in fact, need a brief description, and one reader even suggested to expand the presentation. The present account of development for usability is not linked well enough to the book's theoretical apparatus (cf. Points (1) and (3)), and suffers from a mismatch between being too specific and too general. The new book plan has been designed very much with these problems in mind, cf. Section 6.

Points **(7)**, **(8)**, **(9)** and **(10)** all address presentation issues and will be taken care of in the final manuscript.

3.3 Suggested Additions

1. **Examples**: describe usability techniques applied to the specific multimodal SIMILAR examples when ready.
2. **Standards**: add useful references.
3. **Literature**: more on specific SIMILAR (partner) contributions to the field. The following examples of literature were proposed:
    a. F. Paternò et al., such as "Remote Web Usability Evaluation Exploiting MultiModal Information on User Behaviour" (2006), "Increasing Usability when

Interacting through Screen Readers" (2004), "Models for Universal Usability" (2003), "The Impact of Different Media on Safety and Usability of Interactive ATC Applications" (1999);

b. J. Vanderdonckt et al., such as "Detecting Interaction Variables in a Mixed Reality System for Maxillofacial-Guided Surgery" (2005), "Continuity as Usability Property" (2003);
c. L. Nigay et al., such as "Formal Testing of Multimodal Interactive System" (2006), "Four Easy Pieces for Assessing the Usability of Multimodal Interaction: the CARE Properties" (1995);
d. D. Tzovaras et al., such as "Design and Implementation of Haptic Virtual Environments for the Training of Visually Impaired" (2004)
e. and works by other authors, for example, C. Bach and D. Scapin "Obstacles and Perspectives for Evaluating Mixed Reality Systems Usability" (2004) probably deserve more attention.
f. http://portal.etsi.org/docbox/EC_Files/EC_files/eg_202191v010101p.pdf - the ETSI Guide on "Human Factors (HF): Multimodal interaction, communication and navigation guidelines" (ETSI EG 202 191 V1.1.1 (2003-08)) considers the key issues, solutions and advances in ICT developments in general and multimodal systems in particular. The document is mainly devoted to technologies and multimodal systems for the disabled.
g. http://www.usernomics.com/usability.html - the website of a professional consulting company specializing in user interface design, usability, human factors, and ergonomics offers a large resource in the field.

4. A **chapter** on the basics of how different modalities interact with the human physiology and some physiological basics.

Point **(1)** will be addressed by making substantial use of material from the two usability tests of SIMILAR prototypes, cf. Section 5.

Point **(2)** will be addressed by adding references to relevant bodies for standardisation.

Point **(3)** will be kept in mind when revising the manuscript and references will be added as appropriate.

As for Point **(4)**, although potentially very interesting, this proposed addition would probably be better addressed in a context which focuses more on ergonomics and ergonomic aspects of modality theory and HCI than the present book does.

4 Reviewers' Comments

The two sets of book reviews received in the summer of 2007 provide, first of all, important suggestions on the structure and logic of the manuscript.

Like our colleagues in SIMILAR, the reviewers ask for further clarification of our theory about usability factors and for a stronger bridging between the theoretical chapters of the book (Chapters 3 and 4) and the subsequent chapters, cf. Section 3. In addition, the reviewers very sensibly point out that since the primary novelty of the book is the multimodal aspect, this should be introduced from the start in Chapter 2 followed by the theory of usability factors in Chapter 3. As for the bridging aspect, and in particular with respect to bridging from modality theory to usability development and evaluation practice, the reviewers ask that we spell out more explicitly through analysis, description, and illustration - the *changes* which multimodality brings to traditional GUI-oriented HCI, for instance in terms of new methodologies required for working with users or for analysing user test data, in terms of de-

emphasising traditional methodologies, or in terms of implications for analysis, specification and design.

Regarding the chapters on practical methodology, in particular Chapters 6 through 9, the reviewers ask us to strengthen the meta-discussions about the methodologies in order to provide more useful information on when to use which method or technique, and why.

Finally, the reviewers also comment on various presentation issues, proposing a presentation method of gradual descent into more detail and increased use of graphs and tables, and suggested to use more of the work done in SIMILAR in the final book.

In total, the reviewers made 37 specific comments and suggestions, and we agree with virtually all of them.

5 Lessons Learned from Usability Tests 2007

In June, 2007 the authors conducted two usability tests of SIMILAR-developed research prototype multimodal systems. The purpose of the tests were two-fold: (1) to test new multimodal technologies developed by partners in SIMILAR, and (2) to test the first version of the authors' book manuscript *Multimodal Usability*. For purpose (1) see [Bernsen and Dybkjær 2007]. We now describe work related to Purpose (2).

The systems tested were the following:

1. A speech and 3D gesture input system for playing Sudoku developed at ZGDV in Darmstadt, Germany. The technology used, but not the more recent, actual system tested, is described in [Malerczyk et al. 2005].
2. A strongly multimodal computer game system which enables a blind and a deaf person to collaborate in a treasure quest. This system was developed at ITI-CERTH in Thessaloniki, Greece, cf. [Moustakas et al. 2006].

The two usability tests were carried out with 12 and 6 users, respectively. The Sudoku system was tested in NISLab's usability laboratory and the treasure hunt system was tested with blind users at the Institute for The Blind in Copenhagen.

In preparing and carrying out the usability tests, as well as in validating the collected data, we followed the manuscript's recommendations throughout as documented in [Bernsen and Dybkjær 2007]. Since the authors had done very many usability tests before and on many different multimodal systems, these processes were largely completed as planned. Nevertheless, we learned two major lessons from the tests, one from each test, as we shall now describe.

5.1 Methodology of Analysis: Lessons from the Sudoku System Test

When development and evaluation for usability moves from Graphical User Interfaces (GUIs) to other multimodal interactive systems, two major implications are that (i) new usability *development methodologies* are needed and (ii) a new focus enters, which is that of *modality appropriateness*. Concerning this latter point, since the GUI modalities can no longer be taken for granted, a *key development question* becomes: which modalities to choose for the application, task, intended users, etc., and a *key evaluation question* becomes: are the modalities that were actually chosen the "right" ones or are there any better modality combinations?

The two "backbones" of the draft manuscript are, in fact, meant to support designers, developers, and evaluators in answering exactly those two key questions of modality choice and modality choice evaluation. In the Sudoku system evaluation in [Bernsen and Dybkjær

2007], we have followed this lead and analysed our observations of the subjects while playing the game, as well as their answers to the post-trial structured interviews, as follows:

We first (1) analyse the technical problems identified because that's mandatory for the interpretation of any user response to the system being tested. We then (2) analyse the modalities chosen in-depth to identify as many modality choice issues as possible, including the cases there might be for alternative and/or additional interactive modalities. Thirdly, (3), we analyse the multimodal combination as a whole as to how it works during gameplay. Then we analyse (4) information adequacy *given* the modalities chosen for the game, i.e., given the chosen input/output modalities, should more, less or other information have been exchanged between user and system. Finally, (5), we analyse the functional issues with the system identified through the analysis carried out in Points (2) through (4).

This methodology of analysis, we argue, (i) works well for novel multimodal systems, such as the Sudoku system tested; (ii) is focused throughout upon what is novel to new multimodal systems compared to GUI-based systems; and (iii) is perfectly suited for implementing the "bridge" from theory to practice that is missing in the draft manuscript.

Though less rigidly, we follow a similar approach in the analysis of the usability test results from the treasure hunt game user test in [Bernsen and Dybkjær 2007].

5.2 Opening a New Tactile World: Lessons from the Treasure Hunt System Test

Since the treasure hunt game user test results were analysed after the Sudoku game results, this was not an occasion for discovering major novelties in methodology of analysis. However, the treasure hunt game user *test itself* taught important lessons concerning *how novel and unfamiliar* a new modality combination can be to both subjects and usability testers alike.

Prior to the test, we knew that the user population consisting of blind users were new to us, that we had not had occasion to try out the system much ourselves, and that the multimodal technology probably would be as new to the users as it was to ourselves. We therefore planned a test with few users, designed it to be exploratory rather than focused on gathering data on more or less familiar issues. We also knew that the technology would be at a very early stage in usability terms, development having been focused far more on multimodal technology than on usability. However, despite all of these "keep an open mind" mementos, the test was full of surprises as explained in detail in [Bernsen and Dybkjær 2007]. Here are some of them, in brief.

We needed *more time* for each user than planned and available because the large majority of users had substantial difficulties getting through the game on their own, as a result of which we added an ever-increasing amount of instruction to each new user during gameplay.

The *metaphor* we used in introducing the technology to the users was very likely flawed and tended to hamper their gameplay rather than further it, as a result of which we kept revising the metaphor and our pre-trial instructions.

Within the small 6-person subject population, *variation* was very substantial as regards subjects' ease and speed of gaming, apparently as a function of at least two factors, i.e., computer gaming experience and ease of grasping the nature of gaming with the 3D haptic input/output device used.

At the same time, all users agreed that the game provided them a fantastic glimpse of future technology which could benefit the blind socially as well as opening up a world of combined 3D haptics output and speech output for which an apt analogy is the 3D world as accessed by the seeing.

These lessons will definitely have important implications for the practical advice provided in the final book.

6 Book Revision Plan

Based on the input to the book writing process described in Sections 3, 4, and 5 above, we have revised the plan for the final book as follows.

In a general outline, we propose to do the following – using boldface to highlight the "thread of integration" we would like to follow:

- The **(1) two purposes of the book** are to (1) present the multimodal challenge to usability and its implications for usability methods and techniques, and (2) present central methods and techniques for working with usability in a multimodal context.
- Launch the multimodal usability theme in Chapter 1 and focus the theme on a limited number of **(2) central multimodal challenges** facing developers. "Centrality" is defined by particular combinations of modalities and families of modalities of maximum import, and the issues to be addressed will concern modality selection for applications and usability methods and techniques that are different from those used in developing graphical user interfaces (GUIs) or familiar methods that require revisions to be applied beyond GUIs. The rationale for focusing on "selected multimodal chunks" is that we cannot have umpteen sections each describing a particular modality and what it's good or not so good for, comparing all modalities pairwise for what they are good or less good for, and describing each possible modality combination and what *it* is good or less good for. What we can do is to focus on a limited number of areas which account for most of the novel issues facing multimodal developers today. Two **examples** of challenges are: (i) how do we develop and evaluate usable systems which enable spontaneous and more or less unconstrained spoken and written natural language interaction? (ii) How do we develop and evaluate usable interactive systems that can see?
- Remove Chapter 2 on software engineering. Some remarks on software engineering that we find essential will be included in Chapters 4 and 5 on development and evaluation for usability (see below).
- Present a fully integrated framework for multimodal usability in Chapters 2 and 3 by:
    - not just presenting but **applying modality theory to the challenges** mentioned above, including consideration of the kind of work represented by CARE and Tycoon. This should also make it natural to discuss links with task-level design and some of the typical dilemmas of modality choice. And
    - making it clear that the usability properties are factors influencing usability which must be taken into account throughout the software lifecycle of development and evaluation.
- **Bridging from theory** and discussion of application issues in Chapters 2 and 3 **to concrete development practice** in Chapters 4 and 5 by demonstrating how to use the framework in development (Chapter 4) and how to use it in evaluation (Chapter 5). Each of Chapters 4 and 5 will include a multimodal case study which applies the framework – we have just done such a study in SIMILAR in which the framework was applied to – we believe – good effect. In the context of Chapters 4 and 5, we will address requirements elicitation, including customer interviews.
- Use the multimodal challenges identified in Chapter 1 throughout Chapters 6 through 10 to **orientate all presentation and discussion of methods, techniques and other**

approaches towards the needs raised by those challenges and how those needs are reflected in new methods which are not being applied to GUIs as well as revisions of methods which are also being applied to GUIs. Techniques which are almost uniquely used for GUI development and evaluation should go or be de-emphasised. We fully agree with the importance of **meta-information on methods and techniques and their interrelationships** and will try to provide it in as useful form as we can.

- **Presentationwise**, we will follow the progressive detail-style and also seek to highlight the book's structure better than now.
- As a by-product of the above plan, the **current Chapters 2 and 5 will disappear** and **multimodality will be key from the start** of the book.

In conclusion, it seems to us that the plan outlined above (a) still makes it **feasible** for us to write the book within a fixed time schedule; and (b) **meets all points** raised by the reviewers except two. These are (1) the point made by one reviewer that the book should not be about *multimodal* usability, and (2) the possible point that we should *not try to write a stand-alone book* which on its own can help developers of multimodal systems do usability work knowing what they are up to. We believe that the book will be more useful if the key methods and techniques are there rather than in other books which the reader will then also have to get hold of to work on usability. Having said that, there is, of course, no such thing as a stand-alone book in this enormous field, it's all a matter of more or less, the distinction being, rather, the practical one between a book you can make do with for a start, and a book which has been designed in such a way that you must use other books as well – even for a start.

7 Acknowledgements

The work described in this report was partly funded by the EU Network of Excellence SIMILAR (www.similar.cc). We are grateful for the support.

We would like to thank all colleagues in SIMILAR who took the time to read the book or parts of it and send us their comments: Ana C. Andrés del Valle, Suzanne Kieffer, Pablo Lamata, Andrey Ronzhin, Dimitrios Tzovaras, Jean Vanderdonckt, Jörg Voskamp, and Olga Vybornova.

We are also grateful to the book reviewers for their many important comments and suggestions on how the manuscript can be improved.

Finally, thanks are due to Kostas Moustakas, ITI-CERTH, for taking part in the test of the treasure hunt system in Denmark and to Cornelius Malerczyk for providing the Sudoku system for testing and answering our questions about technical details.

8 References

Bernsen, N. O. and Dybkjær, L.: *Report on Iterative Testing of Multimodal Usability and Evaluation Guide*. SIMILAR Deliverable D98, September 2007.

Bernsen, N. O. and Dybkjær, L.: *Multimodal Usability*. To appear.

Malerczyk, C., Dähne, P. and Schnaider, M.: Exploring Digitized Artworks by Pointing Posture Recognition. *Proceedings of 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, Pisa, Italy, 8 – 11 November 2005.

Moustakas, K., Nikolakis, G., Tzovaras, D., Deville, B., Marras, I. and Pavlek, J.: Multimodal Tools and Interfaces for the Intercommunication between Visually Impaired and "Deaf and

Mute" People. *eNTERFACE'06*, July 17th – August 11th, Dubrovnik, Croatia, *Final Project Report*, 2006.