Copenhagen Business School
HANDELSHØJSKOLEN

Cultural Usability and Human Work Interaction Design – techniques that connects:

Proceedings from NordiCHI 2008 Workshop

Sunday October 19, 2008

Organizers: Torkil Clemmensen, Rikke Ørngreen, Kerstin Roese, Annelise Mark Pejtersen, Lynne Dunckley, Pedro Campos, Xianghong Sun, Pradeep Yammiyavar

Edited by: Torkil Clemmensen

Proposal for Workshop for NordiChi 2008

Torkil Clemmensen, Rikke Ørngreen, Kerstin Roese, Annelise Mark Pejtersen, Lynne Dunckley, Pedro Campos, Xianghong Sun, Pradeep Yammiyavar

Title: Cultural usability and Human Work Interaction Design – techniques that connects

Duration: one full day

Expected number, balance and selection of participants: Between 20 and 40 participants, mainly but not exclusively with a research background. The requirement for participation will be the submission and acceptance of a position paper of max 10 pages in the format accepted for NordiChi. The papers will be peer reviewed for appropriateness to the topic of the workshop.

If you submit a paper, accept that you might be asked to read at least two other papers. And expect to be asked to function as opponent at the workshop to the papers that you have reviewed.

Workshop theme and goals: This workshop analyzes the use of techniques to connect empirical work analysis and interaction design in different cultural contexts. In industry, a wealth of usability evaluation methods is used to evaluate computer software user interfaces and other interactive products: Inspection methods, Workplace observation, Think-Aloud Usability Test, etc. These techniques often give - seemingly - similar results when applied in diverse cultural settings, but experience shows that we need a deep understanding of the cultural, social and organizational context to interpret the results, and to transform it into interaction design.

The workshop will present current research into cultural usability and human work interaction design. Cultural usability is a comprehensive concept which adheres to all kinds of contexts in which humans are involved (private family, work, public and private organizations, nature and climate, etc.).

Topics may include but are not limited to:

- The use of ethnographic methods to generate scenarios and use cases in a foreign country
- Work style modeling in different cultural contexts
- Using card sorting methods to discover cultural differences in information architecture and design preferences
- Usability testing and cultural differences in socializing and communication
- Recruiting 'surrogate users' in usability practice and research
- Socio-technical design in different cultural contexts
- The use of elicitation techniques with different cultures
- Conceptual framework of cultural factors in interaction design
- Cultural aspects of interaction methods and languages
- Evaluation of the cultural effects of new interactive products
- Experiences and best practices of cultural usability

- Localization of usability methods or research methods
- Cultural factors related to usability

Outcomes of the workshop: This workshop analyzes the use of techniques to connect empirical work analysis and interaction design in different cultural contexts. The workshop presentations and discussions may form the input to a call for a special issue.

Relevance to the field: Both in industry and in research there is an interest in understanding cultural issues because there are many cultural factors that influence the outputs of techniques. From an academic viewpoint, the research field of cultural usability and human work interaction design should be analyzed within an expanded diversity of users, contexts, usability techniques and technologies.

Intended audience: Participants will be people attending NordiChi2008 with an interest in connecting empirical work analysis and interaction design in different cultural contexts. In particular we expect participants coming from IFIP WG 13.6 HWID, University of Kaiserslautern USE group, the CultUsab research project members and interested people, the Lab:USE and the HCI Institute at CMU, and more.

Description of activities planned: To provide a good social atmosphere that invites to openness and provides time to reflection, the workshop will be conducted as round table group presentations and discussions facilitated by the organizers, followed by presentations of selected papers in plenum. One or two of the participants may be asked to provide an overview and commentary to sum up all the presentations as the final activity at the workshop.

Needed facilities: Two or three group rooms (one large enough to accommodate 40 people).

Organisers' names and backgrounds:

Torkil Clemmensen, associate professor, Department of Informatics, CBS, Denmark; chair of IFIP WG 13.6 HWID, project coordinator of CultUsab research project

Rikke Ørngreen, Center for Applied Information Communication Technologies, Copenhagen Business School, Denmark; secretary IFIP WG 13.6 HWID and project manager for CaseMaker

Kerstin Roese, User-centered Product Development, University of Kaiserslautern, Department of Mechanical Engineering, Germany

Annelise Mark Pejtersen, professor, Center of Cognitive Systems Engineering and the University of Washington, Denmark and USA. Chair of IFIP TC13 on HCI

Lynne Dunckley, professor, Thames Valley University, London, UK. Director of Research for TVU Institute for Information Technology and Leader of HCI aspects of EPSRC Vesel project

Pedro Campos, assistant professor, Department of Mathematics and Engineering, University of Madeira, Portugal

Xianghong Sun, professor, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

Pradeep Yammiyavar, professor, Department of Design, Indian Institute of Technology, Guwahati, India

Draft programme (subject to change) for Workshop at NordiChi 2008 Sunday October 19:
Cultural Usability and Human Work Interaction Design – techniques that connects

- Organizers: Torkil Clemmensen, Rikke Ørngreen, Kerstin Roese, Annelise Mark Pejtersen, Lynne Dunckley, Pedro Campos, Xianghong Sun, Pradeep Yammiyavar

0900-0930 Socializing

0930 – 1000 INTRODUCTION

- Torkil Clemmensen – introduction to cultural usability and HWID workshop
- Rikke Ørngreen: Themes in Human Work Interaction Design

1000-1140 PAPER PRESENTATION

- Dinesh Katre: One-handed thumb use on smart phones by semi-literate and illiterate users in India
- Huanglingzi Liu: Evaluating Writing Chinese with Finger Tip on Touch Sensitive Keypad
- Qingxin Shi: A Study of Usability Problem Finding in Cross-Cultural Thinking Aloud usability tests
- Lene Nielsen: Different Cultures' Perception of Personas Descriptions
- Pradeep Yammiyavar: Extracting Users' Data: Towards development of a cultural and semantically sensitive combinatorial methodology

1200 – 1300 PANEL DISCUSSION: *Cross Cultural Research Methodology*

- Lynne Dunckley, Xianghong Sun, Ravi Vatrapu, Torkil Clemmensen, Kasper Hornbæk (not confirmed

1300 – 1400 Lunch

1400-1445 CONCEPTUAL FRAMEWORK

- Masaaki Kurosu: Usability and Culture as Two of the Value Criteria for Evaluating the Artifact

1445 – 1625 PAPER PRESENTATION

- Ravi Vatrapu: Cultural Usability in Computer Supported Collaboration
- Xianghong Sun: Do cultural factors affect thinking in usability tests?
- Matthias Rehm: Creating a Standardized Corpus of Multimodal Interactions for Enculturating Conversational Interfaces
- Kerstin Roese: (to be announced)
- Kasper Hornbæk: (to be announced)

1630 – 1730 PANEL DISCUSSION: *Combining usability with empirical studies of human work*

- Masaaki Kurosu, Kerstin Roese, Dinesh Katre, Rikke Ørngreen, Annelise Mark Pejtersen

1730 – 1745 The end, and future activities in HWID and Cultural Usability, Torkil Clemmensen

18.00-20.00 NordiCHI 2008 Reception and Registration

Themes in Human Work Interaction Design

Orngreen,Rikke[1]; Mark Pejtersen, Annelise[2]; Clemmensen, [3]Torkil

[1] Danish School of Education, Aarhus University, Denmark, orngreen@cbs.dk

[2] Chair of IFIP TC 13, Denmark, ampcse@mail.dk

[3] Department of Informatics, CBS, Denmark, tc.inf@cbs.dk

Abstract. This paper raises themes that are seen as some of the challenges facing the emerging practice and research field of Human Work Interaction Design. The paper has its offset in the discussions and writings that have been dominant within the IFIP Working Group on Human Work Interaction Design (name HWID) through the last two and half years since the commencement of this Working Group. The paper thus provides an introduction to the theory and empirical evidence that lie behind the combination of empirical work studies and interaction design. It also recommends key topics for future research in Human Work Interaction Design.

Keywords: work analysis, interaction design, HCI tools, new ways of working

1. Introduction – Scope and Research Area

Technology is changing human life and work contexts in numerous ways: interfaces between collaborating individuals in advanced ICT networks, small and large-scale distributed systems, multimedia and embedded technologies, mobile technologies, and advanced "intelligent" robots. With this change towards new ways of working, an intensive demand has taken place for techniques and technologies that address contemporary issues related to communication, collaboration, learning, problem solving and information seeking in large information spaces of great variability. To address this comprehensive problem, an IFIP working group on Human Work Interaction Design (HWID) was established. Its expressed purpose was to reach a better understanding of the new challenges entailed in the design of technological support for modern, dynamic and complex work environments through a discussion of the interrelation between Work Analysis and Interaction Design within the field of Human Computer Interaction .

The main problem addressed is how we can understand, conceptualize and design for the complex and emergent contexts in which human life and work are now embroiled. This issue calls for cross disciplinary, empirical and theoretical approaches that focus on Human – Work Interaction design.

The main target of this paper is to draw attention to this problem by discussing recent research topics which address this problem using different approaches, and secondly, to point to problems which need to be investigated further. Hopefully, this will encourage more empirical studies and conceptualisations of the interaction among humans, their work and other variegated contexts and the technology used both within and across these contexts.

2. Background

HWID organises Annual Working Conferences with printed papers, discussions and varied forms of interactions and collaborations during two days of workshop activities. HWID'05, a Working Conference took place in Rome at INTERACT'05, the International Conference of Human Computer Interaction. The focus was on "Describing Users in Context". HWID'06 took place at the University of Maidera. The theme was "Synthesizing work analysis and design sketching", with a particular focus on how to read design sketches within different approaches. HWID'07 took place in Rio de Janeiro at INTERACT'07. The focus was on "Social, Organisational and Cultural aspects of Human-Work Interaction Design". The inspiration of this paper is based on fruitful dialogs during these HWID activities.

The subjects raised in this paper stem from the authors' analysis of papers written in conjunction with discussions that took place. These papers were written by researchers from around the world; the topics covered a variety of disciplines and theoretical approaches in human sciences: psychology, anthropology, sociology, information and media sciences; computer sciences and engineering.

Human-Work Interaction Design is a comprehensive approach in HCI, and in order to provide an easy understanding and to illustrate the coverage of this research topic, we developed the model in figure 1.

Figure 1 shows examples of the characteristics of humans and work domain contents and the interaction during their tasks and decision activities, individually or in collaboration. Analysis of users' work and life, as well as the design of computer-based information systems, has inspired the development of numerous theories, concepts, techniques and methods. Some have been widely adopted by practitioners; others are used mainly by researchers, and these are naturally part of Human-Work Interaction design research, and they will obviously influence the work and user analysis as well as the technology design. This explains the top box.

Environmental contexts, such as national, cultural, social and organizational factors, impact the way in which users interact with computers in their work and life to the same extent as the nature of the application domain, the tasks, and the users' skills and knowledge. The analysis and design of Human-Work Interaction will necessarily also include these contextual factors. As a consequence hereof, the bottom box of figure 1.
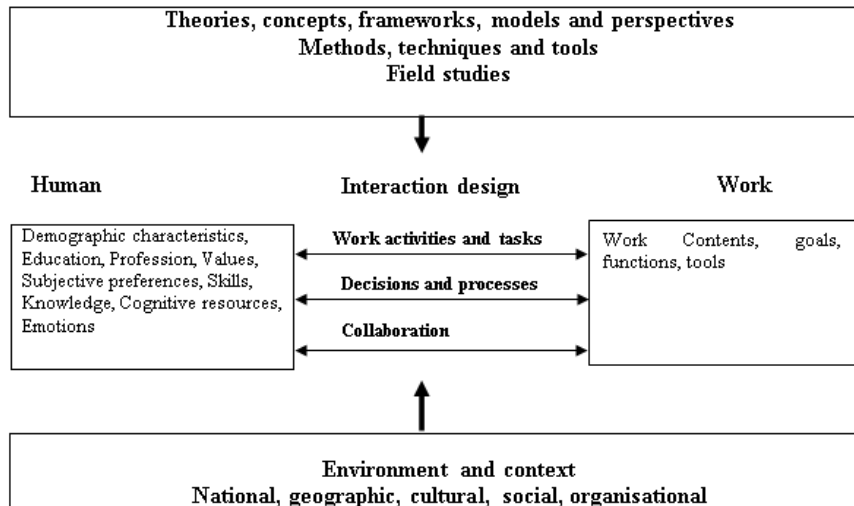
Figure 1 The Model of Human-Work Interaction Design

The analysis of the current HWID activities resulted in the identification of six main themes, which reflect those problems which the authors perceived to be of major concern in Human-Work Interaction Design:

Within Design processes
− Encouraging the dialogue between users and designers in the design process
− Bridging the HCI and Software Engineering gap by working with user requirements and collaboration in software development processes
− Supporting communication and design exploration through sketching

Within Work and User analysis
− Bridging the work analysis and interaction design gap through detailed case and field studies and action research projects empirical field studies.
− Rich contextual user descriptions, including methods to study unpredictable and opportunistic tasks
− Broadening the scope to Social, Organizational and Cultural aspects

Although this list presents these themes and problems as separated, they are of course intertwined and appear in different ways in many of the papers. Thus the following presentation of the papers within one of these headlines is of course exclusive for practical reasons, but as the reader will recognize, there are many overlapping themes and problems.

3. Dialogue in the design process, between users and designers

Design conceptualized as dialogue. Lopes 2006 provide a perspective on design as dialogue, consisting of a presentation of different definitions and

different aspects of design, which could be argued as all being related to dialogue. Dialogue is considered in relation to objects, processes and disciplines of design. The author identifies some problems with the design-as-dialogue approach, mainly the complexity of the issue, and suggests a qualitative study that may help reveal ways to simplify and validate the approach.

Grounded theory to study users' responses. Nocera et al. 2005 suggest ways to support people's meetings and dialogues about their view of the world and their experience. They use grounded theory in the study of users' responses to an implementation of an ERP system in various countries; the authors investigate negotiation – as reconfiguration – between the roles of users and producers. The analysis shows very different attitudes toward the same systems when implemented in cultural diverse settings; it purports that making sense of the system in a particular work-context depends on cultural, organisational and individual preferences. These different attitudes and ways of use are particularly visible in breakdown situations; the authors argue for interaction between users and producers, and that producers should be able to observe and discuss users' breakdown situations, their frustration and workarounds.

Affinity Diagram for requirements elicitation. Bondarenko and Janssen 2005 use a different methodological approach. They use the Affinity Diagram method adapted from Hackos and Redish in the requirements elicitation process for the design of personal document management systems. Without losing the user's context and without requiring the reading of lengthy reports, this method helps structure large collections of mixed qualitative and quantitative data, and gives dynamic requirements (as opposed to static user profiles or task flows). However, the method as it is used per se results in abstraction of the requirements into a general level and hence results in difficulties in mapping the acquired results into system design.

Information acquisition using colleagues' verbal reports. Erlandsson and Jansson 2007. A new method for information acquisition called *collegial verbalisation* is explored using an empirical case study of vehicle operators being videotaped while driving a high-speed ferry, followed by some of their colleagues making verbal reports while watching this video data. These colleagues are very familiar with the driving task and the driver environment. The method is discussed in relation to the amount of information provided in general; the reliability of the data; and how it contributes to the detection of "buggy mental models" within the operators, and it is compared to more traditional forms of verbal reports. It is suggested that the method of collegial verbalisation may have combinatorial advantages that makes it more powerful as an analysis tool than the traditional forms of verbalisation, specifically if one wants to analyse work tasks that are dynamic and where the operators' behaviours are highly automated. However, more elaborate and systematic investigations must be conducted through experimental designs.

4 Bridging the HCI and Software Engineering gap

User interface model and requirement tool. España et al. 2006 look at the gap between HCI and Software Engineering (SE); while SE is supposed to be strong in specifying functional requirements, HCI is centred on defining user interaction at the appropriate level of abstraction. An abstract model of the user interface represented by a ConcurTaskTrees model is used to enrich the functional specification, and a new tool called RETO that aims at requirement engineering is presented. The adoption of such a framework is promising and future empirical studies will show if the model can be justified.

Activity Theory and software development process. Software development is intrinsically a collaborative activity. Based on an analysis of current literature and software, Lewandowski and Bourguin 2006 find that current Software Development Environments seldom provide true integrated collaboration between developers, rather they offer only sharing of material or communication support, and do not support the actual work process of software development. Further, the ability to tailor the development environment is an issue, as features for allowing external applications to be nested into the environment are lacking. Grounding their work on Activity Theory, the authors describe how the eclipse (open source software) has been extended to accommodate for some of these deficiencies; it will be interesting to follow these features being implemented in future development processes.

User interface patternsin specific contexts. Stanard and Wampler 2005 focus on richness multi-dimensionality of user descriptions, and discuss how design patterns until now have been close to traditional usability guidelines; thus, there is a need to make design patterns to better support interaction of specific contexts. User Interface (UI) patterns are presented as a way of defining, applying and evaluating the translation of cognitive and collaborative requirements into meaningful human computer interaction in the designed interface, and then through this provide input to the development process. The described case involves an airport control system, and the discussion of command and control systems. The patterns are useful not only for the provision of training and inspiration to solutions, but also for the reuse of patterns that have been quality assured in complex and risk environments, such as command and control systems. The authors argue for the need for hierarchies of patterns that are based on a specific application-domain or work-domain to enhance the work-performance.

Work style modelling In the same vein, Campos and Nunes 2005, 2006 combine Work Style modelling with Usage-Centered Design with the objective of designing and evaluating better design tools. They describe the richness in the human-work interaction by using a new method of work style modelling, which has been applied to the work-context of interaction designers (as well as to collaborative software design). The work style is described from a set of informally defined values, and the set of styles which has been shown apparent in the work-context, are then more formally depicted and evaluated using diagrams

and metrics. By modelling users' work style, the focus is put on work transitions (from one style of work to the other) and the designed solution ability to support the current context and changes in these – within the same application. The authors raise the question of whether it is possible to use work style modelling in other fields to describe flows between contexts of use.

5 Sketching in Communication and design exploration

Collaborative design process. Craft and Cairns 2006 offer experiences with sketching in a design process for an information visualization tool. The objective of the system is to support communication between users with different backgrounds - between biologists and mathematicians. The authors present an in-depth analysis of the design process, showing that sketching as an integral part of a collaborative design process aids creativity, communication, and collaboration.

Representation of requirements based on Cognitive task analysis. Rozzi and Wong 2006 present a case study of how design sketching can be used as a technique for the representation of design requirements to help the creation of a common understanding between users, designers and software developers, during the development of a tool for supporting spatial-temporal reasoning in Air Traffic Control (ATC). The design process is based on a cognitive task analysis using the Critical Decision Method, relying on observation and video recordings as well as Contextual Inquiry interviews. The authors show how sketching was used to get insights into the design possibilities, but also find that spatial-temporal issues are difficult to illustrate with sketching techniques; thus, further work is needed.

Idea exploration and refinement of details. Orngreen 2006 reflects on what sketches are and on the use of design sketches when developing an e-learning platform for case-based learning. The author attempts to differentiate techniques that include sketches: rough hand drawn sketches - storyboards – prototypes, and how the emphasis changes from idea exploration to refinement of detail. The paper draws a distinction between a sketch as a design artefact that can stand alone and as part of a work process.

Reading design sketches using work analysis. Clemmensen 2006 investigates the role of design sketches in Interaction design and work analysis when designing a simple folder structure for e-learning software to be used for course administration at a higher education study programme. The author discusses how to conceptualize the process of reading design sketches using work analysis. The interface was evaluated using a think-aloud protocol, and was found to be less satisfactory than the earlier designs as it was 'long-winded'. This pointing to the need for future work on investigating the relation between the sketching techniques used and the design obtained in the development process.

Sketches to improve task performance. Although Pereira et al 2006 do not act in the space of IT, they adopt a human centred approach, illustrated with

sketches, when looking to improve the performance of treadle pumps, to be used in developed countries. Similarly, Gaspar et al 2006, use annotations (words and sketches) to the photographs in their analysis and design studies when investigating ways to increase the amount of physical activity in the daily routine.

6 Bridging the gap between work analysis and interaction design

Cognitive work analysis and interface design. Upton and Doherty 2006 describe an approach to designing a visual application for a semiconductor manufacturing plant, which is seen as a complex, large-scale system requiring a structured design methodology. They present a design rationale supporting the explicit representation of hierarchies, the compatibility of views, and the use of contextual navigation. This design is derived from a cognitive work analysis, from which an Abstraction Decomposition Space (ADS) was made and the interface design was subsequently developed. The paper systematically describes the application of cognitive work analysis and the subsequent process of interface design, in an effort to bridge the design gap.

Future/vision seminars in action research. Based in user-centred and participatory design, Johansson and Sandblad 2006 investigate how a home care and help service organisation can be developed in order to be better prepared for future challenges. During their action research project, they used the future/vision seminar model, extended with assignments (such as: describe a day at work). The seminars resulted in the formulation of several scenarios, which again served as input to the design of a prototype.

Generic user interface for resource allocation. O'hargan and Guerlain 2006 provide a generic User Interface (UI) design for resource allocation problems. The UI is designed to support a person making resource allocation decisions (as opposed to purely automated decisions, often currently the case). They argue that their Resource Allocation Planning System (RAPS) can be adapted to several types of resource allocation domains. In future work it will be interesting to follow evaluations on whether or not it is capable of clearly supporting the work of people doing resource allocation.

Cognitive Task Analysis and Mapping analysis of team performance. Mapping analysis results into new designs in a multi-agent world. This is the focus of the proposal by McMorrow et al. 2005, who use cognitive task analysis to evaluate effective team performance in collaborative environments, such as air traffic management, in order to provide insights into how a technology becomes a 'team player'. A cognitive task analysis for effective team performance can help re-interpret the formal procedures often surrounding complex technological designs by negotiating among different perspectives and different meanings brought into the work environment.

Cognitive Work Analysis and train driver interfaces. Jansson, Olsson and Erlandsson 2007 conducted field studies on the improvement of existing train driver interfaces within the framework of cognitive work analysis (CWA) (Rasmussen, Pejtersen, Goodstein 1994 together with the method for collegial verbalisation which produces think-aloud protocols from video-recordings. The analyses show that the driver works in three rather separate time intervals: a long-range, a short-term and an immediate sense perspective. The driver switches between these while travelling between two stations. A prototype of a planning area of a driver interface was developed, making these switches and feed-forward planning possible. Early tests using the user centred design approach show that the planning area of the interface supports the feed-forward decision strategy. However, the driver group also made substantial changes in the design, indicating that UCSD is an efficient tool in order to capture user competencies, and to bridge the gap between analysis and design.

7. Rich contextual analysis of users

Multidimensional, multimedia portraits of users. Recognizing the need for a general format for user descriptions, Orngreen et al. 2005 present a theoretical focus on human beings as they are perceived by the designers of the technologies of the 21st century. They argue that today software developers use techniques and methods in software development that embed mono-cultural and mono-dimensional models in various contexts which in the future must be replaced by rich portraits of human beings. In continuation thereof, the same group of authors in Nielsen et al. 2006 argues that cultural embeddings are significant in relation to HCI because the cultural context is also embedded in the methodological framework, the techniques and the tools that we apply. The authors suggest a research program that aims at developing a theoretical framework supporting the creation of rich multimedia portraits of the human user of multimodal technologies Orngreen et al. 2005; the authors point to a theory of complementary positions that insists on solid accounts from all observer positions in relation to perspective, standpoint and focus Nielsen et al. 2006.

Activity theory, situated action and distributed cognition models. The need for different positions is also a theme in Kimani et al. 2005 who use activity theory, situated action and distributed cognition models to study the nature of tasks in real world, natural settings. Within the context of mobile computing, they focus on how supplementary tasks, such as interacting with the device, are performed while the user does another primary task. Unpredictable and opportunistic tasks can be studied with these beyond task-centric approaches in order to provide rich and complex descriptions of users in the mobile domain. Information Science is another domain, which requires discussion of current approaches to model and describe empirically the different kinds of contexts.

Information science Pejtersen et al. 2005 purport that we need not only an analysis of users' perceptual, cognitive, and social states, but also a deep understanding of how the users' contexts influence their interaction with artefacts such as a Digital Library. They propose that the problems raised within the information science field can provide a number of useful issues for discussion of the current approaches to describing users in context within the HCI field.

Critical Decision Method, Ethnography and Cognitive Work Analysis. Ham et al. 2005 present three case studies using three different methods, two for task-oriented design contexts (the Critical Decision Method and the Ethnography Method) and one for functional-oriented design contexts (the Cognitive Work Analysis Method, in particular the Abstraction Hierarchy). They argue that the critical decision method and the ethnography method provide useful and effective descriptions, enabling task-based design requirements in contexts of anticipated situations, while the abstraction hierarchy provides useful and effective descriptions in work domains of revolutionary designs for unanticipated situations. However, they miss an integrated method for obtaining information about user contexts, a method that is both task- and function-oriented.

The Activity Interview and Activity theory in HCI. Duignan, Noble and Biddle 2006 elaborate on their work on the activity interview based on cultural historical activity theory and in particular the activity checklist. The activity interview uses questions to get to an activity analysis as opposed to the abstract formulations of the activity checklist. The paper gives a thorough view of the activity theory relation to the HCI field and the activity list, and provides critical reflection of the list based on previous literature, as well as on personal experience. These discussions clearly bring forward issues for improvement at a very concrete level. In the future it will be interesting to follow the consequences that the activity interview has on design suggestions and how it can be seen in the resulting design. Further, it will be noteworthy to see whether the interview, as claimed, is appropriate for guiding the process of activity analysis, if performed by those who do not know activity theory or cultural historical activity theory.

8. Impact of social, organizational, cultural and historical factors

Avoiding cultural bias in usability tests. Clemmensen 2007 The CULTUSAB project is conducting an in-depth investigation of the key dimensions of culture that affect usability testing situations, including language, power distance, and cognitive style. All phases of the usability test are being evaluated for cultural impact, including planning, conducting, and reporting results. Special attention is being focused on subject-evaluator communication and cultural bias in the test design and structure of the user interface being tested. Experiments are being replicated in three countries: Denmark, India and China. The research will result

in new testing methods and guidelines that increase the validity, by avoiding cultural bias, and allow for production of comparable results across countries.

Historical, national, and cultural factors in the work place. Rasmussen 2007 presents an empirical, qualitative study of Internet use in a National Film Archive in an Eastern European country. The purpose was to identify the use of and the attitude towards the Internet through field studies of individuals and organizations. The empirical study shows, that the staff at the archive only uses the Internet moderately in their work. It also shows that historical, national, and cultural factors can be used to explain the way people at work reacts to the new Internet technology. A cross-disciplinary study of the literature about Central and Eastern Europe made it possible to explain their behaviour and attitudes within a broader context.

A game based on cultural common sense. Anacleto Coutinho et al. 2007 argue that an effective educational process has to be instantiated in the local culture and that common sense knowledge represents culture. Common sense based games can be used to work on topics taught by teacher and can promote a meaningful learning, since the new knowledge (formal knowledge presented during classes) is related to pieces of knowledge already in the learners' cognitive structure (common sense knowledge). A common sense based game prototype to support the process of knowledge reinforcement of the content presented to students is presented. Teachers can set up a quiz game based on the Brazilian common sense knowledge. Preliminary analyses with users point out the potential for such approach.

9. What did we learn?

Obviously, a long list of specific and important problems can be derived from this research as described in each paper above. However, common issues are also addressed, which concern basic conditions of the HCI research.

While certain techniques and methods provide an integrated focus on analysis *and* design, most focus on either analysis or design. The strongest link between analysis and design is the general reliance on iteration as a way of developing products that fit the user needs and context, but within HWID other means and techniques have also been applied. Our papers and activities in the Working Group have operated on three levels:

- A field study level which involves an understanding of what actually goes on in a user environment.
- An applied level, which concentrates on methods and tools for analysis and design
- A theoretical level where academic disciplines have been selected to compensate for the shortcomings of single approaches when confronting the complexity of a design problem.

While experimental design of prototypes is a necessary component of the iterative process of work studies, design and evaluation, consistent conceptualisations between work analysis and application evaluation are needed to provide results that are valid beyond discrete experiments, and can be generalised to other application domains and contexts. In some papers the authors present a satisfactory result of the application of a specific approach to solve their defined problem, few are not successful, but the majority of papers present approaches to their problems which the authors find promising, although still problematic, or yet unresolved, because no evaluation has taken place, or because it is unknown whether the approach can be generalized beyond the application domain.

The diverse combination of the approaches have mostly been driven by a particular work domain context, which is why the concluding discussion of theoretical concepts and tools applied in empirical work and prototype designs often refer to further research for validation of these in other application domain.

It is obvious that further work needs to be done in evaluating the designs that have been made, not only as they work in everyday practice, but also in relating them back to the insights that were gained from the initial work analysis and interaction design phases; in this way it is possible to better inform the concepts, methods and techniques applied.

Figure 2 shows the human actors who interact with work domains during their collaborative tasks and decision activities. A variety of application domains are studied in HWID research papers and the humans who perform this work also spans many different characteristics. Within HWID many means and techniques have been applied to study particular design problems, in most papers not one, but several theories, concepts, techniques and methods from several scientific disciplines have been necessary.
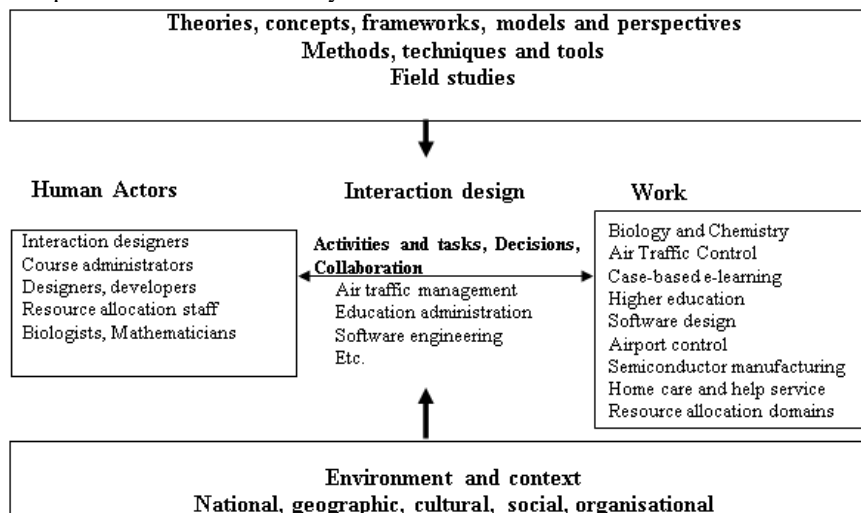

Figure 2 shows the application domains, the tasks and the users involved. The contextual factors in the buttom and the scientific approaches at the top.

There are domains where the work analysis shows that ICT are not the obvious solution by adopting a human centred approach, illustrated with sketches, when looking to improve the performance of treadle pumps, to be used in developed countries (Pereira 2006). Similarly, Gaspar et al 2006, use annotations (words and sketches) to the photographs in their analysis and design studies when investigating ways to increase the amount of physical activity in the daily routine.

Necessary in a global world, but still very emergent area in HCI with few research papers submitted, is to provide a better understanding of the complex interplay between individual, social, organizational, cultural, historical and national factors during the use of technology now and in the future.

10. Conclusion

Finally, we need to mention that although we have introduced many and most, not all, of the hot issues from our papers contributed by participants at the IFIP HWID Working Conferences, this paper's contribution is mostly to give an up to the minute account of research approaches within Human-Work Interaction Design. The informed reader will recognize that apart from the focus on work studies, many of the problems and approaches presented in this sketch are common for other HCI activities, although based on a relative small amount of papers, generalizations are not possible.

In spite of this limitation, it is our hope that the variety of challenges presented in this paper will inspire other researchers and readers to participate and contribute to a better understanding of the complexity involved. We hope this paper promotes the use of knowledge, concepts, methods and techniques that enables work and user studies and design experiments to procure a better apprehension of the complex interplay between individual, social, cultural and organisational contexts.

References

Anacleto Coutinho, Ferreira, Pereira (2007) "Promoting Culture Sensitive Education through a Common Sense Based Game" in Mark-Pejtersen, Clemmensen et al (Eds.) (2007): *Proceedings of the workshop: Social, Organisational and Cultural aspects of Human-Work Interaction Design, at the INTERACT 2007* conference, Rio, Brazil.

Bondarenko & Janssen (2005). "Affinity diagram method: Bringing users' context into the system design", Proceedings of the workshop: *Describing Users in Context – Perspectives on Human-Work Interaction*, the INTERACT 2005, Rome, Italy, p.34-37.

Campos & Nunes (2005). "A human-work interaction design approach by modeling the user's work styles", Proceedings of the workshop: *Describing Users in Context – Perspectives on Human-Work Interaction*, the INTERACT 2005, Rome, Italy, p.44-47.

Campos & Nunes (2006): "Principles and Practice of Work Style Modeling: Sketching Design Tools", Clemmensen, T., Campos, P., Ørngreen, R., Pejtersen, A. M., & Wong,

W. (Eds.). (2006). *Human work interaction design - designing for human work.:* Springer, New York. p. 203-220

Clemmensen (2006): "A simple design for a complex work domain", Clemmensen, T., Campos, P., Ørngreen, R., Pejtersen, A. M., & Wong, W. (Eds.). (2006). *Human work interaction design - designing for human work.*: Springer, New York. p. 221-240

Clemmensen (2007) "The Cultural Usability (CULTUSAB)" in Mark-Pejtersen, Clemmensen et al (Eds.) (2007): *Proceedings of the workshop: Social, Organisational and Cultural aspects of Human-Work Interaction Design, at the INTERACT 2007* conference, Rio, Brazil.

Craft & Cairns (2006): "Using Sketching to Aid the Collaborative Design of Information Visualisation Software", Clemmensen, T., Campos, P., Ørngreen, R., Pejtersen, A. M., & Wong, W. (Eds.). (2006). *Human work interaction design - designing for human work.*: Springer, New York. p. 103-122

Duignan, Noble & Biddle (2006): "Activity Theory for Design: From Checklist to Interview*",* Clemmensen, Campos, Ørngreen, Pejtersen, & Wong (Eds.). (2006). *Human work interaction design - designing for human work.*: Springer, N. Y.. p. 1-26

Erlandsson & Jansson (2007) "Collegial verbalisation – a case study on a new method on information acquisition" in Mark-Pejtersen, Clemmensen et al (Eds.) (2007): *Proceedings of the workshop: Social, Organisational and Cultural aspects of Human-Work Interaction Design, at the INTERACT 2007* conference, Rio, Brazil.

España, Pederiva, Ignacio Panach, Abrahão & Pastor (2006): "Linking requirements specification with interaction design and implementation*",* Clemmensen, T., Campos, P., Ørngreen, R., Pejtersen, A. M., & Wong, W. (Eds.). (2006). *Human work interaction design - designing for human work.:* Springer, New York. p. 123-134

Gaspar, Ventura, Pereira & Santos (2006): "Continuous fitness at home: Designing exercise equipment for the daily routine*"* Clemmensen, T., Campos, P., Ørngreen, R., Pejtersen, A. M., & Wong, W. (Eds.). (2006). *Human work interaction design - designing for human work.:* Springer, New York. p. 147-160

Ham, Wong & Amaldi (2005). "Comparison of three methods for analyzing human work in terms of design approaches", Proceedings of the workshop: *Describing Users in Context – Perspectives on Human-Work Interaction*, the INTERACT 2005, Rome, Italy, p7-11.

Jansson, Olsson & Erlandsson (2007): "Bridging the gap between analysis and design: Improving existing driver interfaces with tools from the framework of cognitive work analysis" in Mark-Pejtersen, Clemmensen et al (Eds.) (2007): *Proceedings of the workshop: Social, Organisational and Cultural aspects of Human-Work Interaction Design, at the INTERACT 2007* conference, Rio, Brazil.

Johansson & Sandblad (2006): "VIHO - Efficient IT Support in Home Care Services*",* Clemmensen, Campos, Ørngreen, Pejtersen, & Wong, (Eds.). (2006). *Human work interaction design - designing for human work.: Springer, New York.* p. 47-66

Kimani, Gabrielli & Catarci (2005)."Designing for primary tasks in mobile computing", Proceedings of the workshop: *Describing Users in Context – Perspectives on Human-Work Interaction*, at the INTERACT 2005, Rome, Italy, p.11-17.

Lewandowski & Bourguin (2006): "Improving collaboration support in software development activities" in Campos, Clemmensen, Orngreen, Wong, Mark-Pejtersen (2006) *the Pre-conference proceedings of HWID'06.* p. 33-45

Lopes (2006): "Design as Dialogue – a New Design Framework", Clemmensen, T., Campos, P., Ørngreen, R., Pejtersen, A. M., & Wong, W. (Eds.). (2006). *Human work interaction design - designing for human work.*: Springer, New York. p. 241-250

Mark-Pejtersen & Fidel (2005). "A Multi-Dimensional Approach to Describing Digital Library Users in Context", Proceedings of the workshop: *Describing Users in Context – Perspectives on Human-Work Interaction*, the INTERACT 2005, Rome, Italy, p.18-26.

Mark-Pejtersen, Clemmensen et al (2007): *Proceedings of the workshop: Social, Organisational and Cultural aspects of Human-Work Interaction Design, at the INTERACT 2007* conference, Rio, Brazil.

McMorrow, Amaldi & Boiardi (2005). "Approaches to designing for highly collaborative, distributed and safety-critical environments", Proceedings of the workshop: *Describing Users in Context – Perspectives on Human-Work Interaction*, the INTERACT 2005, Rome, Italy, p.38-43.

Nielsen, Yssing, Clemmensen, Orngreen, Nielsen, Levinsen (2006):" The human being in the 21st century– Design perspectives on the representation of users in IS*"* Clemmensen, Campos, Ørngreen, Pejtersen, A. M., & Wong, W. (Eds.). (2006). *Human work interaction design - designing for human work.*: Springer, New York. p. 93-102

Nocera, Dunckley & Hall (2005). "Reconfiguring producers and users through human-work interaction", Proceedings of the workshop: *Describing Users in Context – Perspectives on Human-Work Interaction*, at the INTERACT 2005, Rome, Italy, p.27-33.

O'Hargan & Guerlain (2006): Design of a Ressource Allocation Planning System*"* Clemmensen, Campos, Ørngreen, Pejtersen, & Wong, W. (Eds.). (2006). *Human work interaction design - designing for human work.: Springer, New York.* p. 67-92

Orngreen (2006): "The Design Sketching Process", Clemmensen, T., Campos, P., Ørngreen, R., Pejtersen, A. M., & Wong, W. (Eds.). (2006). *Human work interaction design - designing for human work.:* Springer, New York. p. 185-202

Orngreen, Clemmensen, Nielsen, Christiansen, Levinsen, Nielsen & Yssing (2005) "The Human Being in the 21st Century –Design perspectives on the representation of users in IS development, Proceedings of the workshop: *Describing Users in Contex – Perspectives on Human-Work Interaction*, at the INTERACT 2005, Rome, Italy, p.54-59.

Pereira, Malca, Gaspar & Ventura (2006): "Human Motion Analysis in Treadle Pump Devices", Clemmensen, Campos, Ørngreen., Pejtersen, & Wong (Eds.). (2006). *Human work interaction design - designing for human work.: Springer, New York.* p. 135-146

Rasmussen (2007), "Internet Use in Eastern Europe A Case Study" in Mark-Pejtersen, Clemmensen et al (Eds.) (2007): *Proceedings of the workshop: Social, Organisational and Cultural aspects of Human-Work Interaction Design, at the INTERACT 2007* conference, Rio, Brazil.

Rasmussen, Jens, Pejtersen,A.M. and Goodstein, L.P. (1994): Cognitive Systems Engineering. JohnWiley, London.

Rozzi & Wong (2006)" Design Sketching for Space and Time*"*, Clemmensen, T., Campos, P., Ørngreen, R., Pejtersen, A. M., & Wong, W. (Eds.). (2006). *Human work interaction design - designing for human work*.: Springer, New York. p. 161-184

Stanard & Wampler (2005). "Work-centered user interface patterns" Proceedings of the workshop: *Describing Users in Context – Perspectives on Human-Work Interaction*, at the INTERACT 2005, Rome, Italy, p.48-53.

Upton & Doherty: "Visual Representation of Complex Information Structures in High Volume manufacturing", Clemmensen, T., Campos, P., Ørngreen, R., Pejtersen, A. M., & Wong, W. (Eds.). (2006). *Human work interaction design - designing for human work.*: Springer, New York. p. 27-46

One-handed thumb use on smart phones by semi-literate and illiterate users in India: A usability report with design improvements for precision and ease

Dinesh Katre

Group Coordinator

Centre for Development of Advanced
Computing (C-DAC), Pune, India

dinesh@cdac.in

ABSTRACT
There is tremendous potential for developing mobile-based occupational productivity tools for semi-literate and illiterate users in India. One-handed thumb use on the touchscreen of smart phone or touch phone is considered as an effective alternative than the use of stylus or index finger, to free the other hand for supporting occupational activity. In this context, usability research and experimental tests are conducted to understand the role of fine motor control, usability of thumb as the tool for interaction and the ergonomic needs of users. The paper touches upon cultural, racial and anthropometric needs related with the topic, which also need due consideration while designing the mobile interface. Design recommendations are evolved to improve overall effectiveness of one-handed thumb use on smart phone, especially for the benefit of semi-literate and illiterate users.

Keywords
Semi-literate, illiterate, cultural and racial factors, smart phone, touchscreen interface, productivity tools, thumb interaction, stylus, usability, fine motor control, ergonomics

1. INTRODUCTION
1.1 The Indian Context
Presently, India has approximately 286 million mobile subscribers [12] and this number is likely to grow up to 560 million by 2012 [13]. This growth will encompass a significant portion of semi-literate and 400 million illiterate population of India [21]. Telecom infrastructure in India is expanding its reach rapidly [3]. People from all walks of life are beginning to use smart phones and touch phones with touchscreen interface. The cell phone markets in developed nations are getting saturated. As a result, the cell phone makers have to drop prices, and come up with low-cost mobile phones for the markets in less-developed countries like India and China [16]. With the enhanced capabilities and rich features of such mobile devices, it has become possible to use them for a variety of personal and professional tasks [15].

Presently, smart phones are designed to suit the needs of affluent users. But there is a huge opportunity for designing innovative mobile applications for an entirely new segment of users like fishermen, farmers, carpenters, electricians, fabricators, vegetable merchants, shopkeepers, drivers, transport managers, traffic controllers, factory workers and even the housewives in India. Furthermore, many other field occupations can be considered. Mobile based productivity tools and applications can be developed to facilitate the activities like task planning, scheduling, estimation, order booking, accounting, management, etc.

When it comes to spending, there is a difference in the requirements of affluent smart phone users and the potential semi-literate, illiterate users.

Table 1. The requirements of smart phone users	
Affluent Users	**Semi-literate / Illiterate (Potential Users)**
Feature richness	Basic features
Diverse applications	Productivity tools
Self-sufficient device	Helpful in occupational work
Status indicator	Problem solver
Self-indulging	Field attention
Entertainment	Entertainment
Cost is no barrier	Cost effectiveness

Proposed productivity tools for semi-literate and illiterate users should be designed to require minimum text input like numeric data. Most of it would involve selection from a list of pre-defined options, checking, unchecking and iconic buttons. Therefore, the success of proposed mobile-based productivity tools will primarily depend on the usability of interaction tool and user interface design.

1.2 The Choice of Interaction Tool
1.2.1 Unsuitability of Stylus and Index Finger
PDA phones need stylus and iPhone requires index finger for interaction. Our hypothesis is that stylus is more suitable for users with good fine motor control, which is developed out of writing practice. Hand-eye coordination, and dexterity are the other essentials for using the stylus. The semi-literate and illiterate users may not have these skills.

Interaction using stylus and index finger pose similar disadvantage of engaging both hands and attention of user. It constrains our movement and field activity. Coordination between various elements like holding the smart phone in one hand, the other hand with its index finger or the stylus, eyes, attention and fieldwork can be very demanding. This approach has many overheads, which need to be reduced.

Use of stylus or index finger does not seem like a suitable proposition for semi-literate and illiterate users, as they need to do lot of field work and physical activity. Engaging both hands for mobile interaction is against the objective of proposed mobile-based productivity tools. Higher productivity can be achieved by providing a solution that maximizes the use of available resources and minimizes the overheads.

1.2.2 Advantages of One-handed Thumb Use
In this situation, one-handed thumb use on the touchscreen of mobile device seems more suitable to allow the field activity. It can truly help in boosting their productivity. One-handed thumb use on touchscreen can benefit the users by freeing a hand for variety of physical and attentional demands [15]. It will allow the users to hold things or perform some activity using one hand, and simultaneously operate the mobile device using the other hand.

Thumb is well known as the opposing finger, rapid finger and the most dynamic of all fingers. This is already proven by the research on thumb kinematics [17].

1.3 Varying Characteristics of Thumb
1.3.1 Physical, Cultural and Racial Characteristics
Use of stylus insulates a whole lot of physical, cultural and racial issues, which begin to surface while using the thumb for mobile interaction.

Anthropologists have already presented the heterogeneous racial characteristics and differences in the anatomical proportions of human beings belonging to different races [4]. Singh et al. have analysed and shown the effect of habitual activities on body proportions (including the shapes and sizes of hands) of traditional occupational groups [20]. Roughness, softness and cleanliness of hand also depend on one's occupation and work environment.

Anthropometric data clearly indicates varied average heights of males and females from different countries and races. Average heights of Indian male and female are 165.3 cm. Average heights of white American male and female are 178.2 cm and 164.1 cm respectively [24]. Proportionately the size and shape of fingers also change from race to race and country to country. Therefore, the ideal target size proposed [8, 15] for thumb interaction may be biased towards a particular race or culture.

Most semi-literate and illiterate Indian women wear colorful bangles in both hands, which can also affect their interaction with mobile device.

1.3.2 Ergonomic Suitability
A recent study has revealed thumb related Repetitive Stress Injuries (RSI) among the mobile workers who extensively used BlackBerry [22]. In this context, it becomes very important to study the ergonomic aspects of thumb. The RSI continue to persist despite several ergonomic studies to design thumb keyboard [25].

On considering the multiple dimensions of one-handed use of thumb on smart phones by semi-literate and illiterate users, a focused usability study is undertaken with the following objectives.

1.4 Objectives of Usability Study
- Study the feasibility of one-handed thumb use by semi-literate and illiterate users. Understand the role and impact of fine motor control, when they use stylus or thumb on smart phone. Identify their proficiencies and limitations.

- Study the usability, physical characteristics, qualitative, and ergonomic aspects of thumb as the tool for interacting with smart phones. Identify the factors that affect the quality of interaction.

- Evolve design recommendations based on the observations for enhancing the overall effectiveness of interaction using thumb.

(In the current scope of this research, our emphasis is more on collecting the qualitative and observational data.)

2. RELATED WORK
The fundamental study on human motor systems in controlling the amplitude and movement by Fitts [5] has complemented our experimental work. We have adopted some aspects of the reciprocal taping apparatus given by Fitts, as part of the tests performed using thumb and stylus.

Karlson et al. have presented thumb gesture based Scalable User Interface (ScUI) technique to support multiple devices with different resolutions and aspect ratios [8, 9]. Fisheye and pure zoom techniques are used over multiple applications for magnifying the user interface. Thumb gestures also require certain amount of fine motor control. Our study focuses on identifying the role of fine motor control and its availability with semi-literate and illiterate subjects, which is an imperative for touchscreen interaction.

Wu et al. have carried out the performance study on touch-pen sizes in three screen tasks. They used 12 different touch-pens varying in length and diameter to perform pointing, clicking, writing and drawing tasks on the tablet PC [23]. This study has shown that the quality of performance can vary depending on the length and the diameter of touch-pen. We found this research very useful, as it creates a ground for conducting similar research on the use of thumb for interacting with smart phones. Our explorations identify the variable factors of thumb that impact the precision and overall quality of interaction.

Parhi et al. have conducted a study to identify optimal target size for one-handed thumb use on small touch screen devices [15]. It involved experiments on discrete and serial tasks and hit response variability for different sizes of targets and locations. Their study recommends sufficiently large target size of 9.2mm for discrete tasks and 9.6 mm for serial tasks for one-handed thumb use on touch screen handheld devices. These experiments were carried out on 20 right-handed subjects in the age group of 19 to 42 years using HP iPAQ h4155 PDA. These findings may be true for certain class of users only. In our study, the socio-cultural context is Indian and the targeted users are semi-literate and illiterate. Our focus is not on identifying the target size but on evaluating the usability of thumb as an interaction tool. We explore the design enhancements necessary for improving precision and ease of thumb interaction.

Yunfei has carried out an ergonomic experiment for thumb keyboard design [25]. Balakrishnan el al. have investigated the effect of varying sizes of thumbs, while using the mobile keypad for texting [1]. Both of these experiments are complementing to our experiments, as they have considered some of the anatomical aspects of thumb while evaluating the keyboard of mobile phones.

Gokturk et al. have analysed the index finger as a pointing device in comparison with mouse and joystick [6]. Apart from these, there have been several studies to provide tactile feedback for finger based text entry [2], to understand the problems involved in the bilingual keypad layout of mobile handsets [10] and to evolve guidelines for the button sizes [18, 14] for old persons [11, 19].

3. USER CATEGORIES

The categories of users are broadly defined as under.

Semi-literate Users-
- These are such people who have left the education half way due to economic reasons. But they are familiar with reading and writing.
- Their occupation is such that they seldom require to write or they do not require to write anything. Their writing skills have significantly eroded.
- Their occupation involves lot of physical activity and fieldwork.
- They are familiar with mobile phones.

Illiterate Users-
- These are such users, who have never been school. They do not have reading or writing skills.
- They can understand English numbers. They have learnt basic addition and subtraction of numbers through financial transactions, as the values on Indian currency notes and coins are mentioned in English.
- Their fine motor control (specifically the skill that is developed out of writing practice) is not well developed due to lack of education.
- Their occupation involves major physical activities and fieldwork.
- They are aware of mobile phones but they may not have used it.

We focus on such semi-literate and illiterate users, whose occupation / business justifies the use of mobile-based productivity tools in terms of its complexity, revenue, activity, etc. Sample occupations are enlisted earlier in point 1.1. For the purpose of comparison, the characteristics of educated users are also considered.

Educated Users-
- They have completed education till at least graduation level. They are proficient in reading and writing.
- Their occupation is such that they need to read and write quite frequently.
- Their fine motor skills are well cultivated.
- They are very familiar with mobile phones.

Following short forms are introduced while discussing the experimental tests-

Educated Subjects (ES), Semi-literate Subjects (SLS), Illiterate Subjects (IS)

3.1 Participants

Our experiments were performed on total 21 participants within the age group of 25-40 years. The details are as under-

- Semi-literate Subjects (SLS): 7
- Illiterate Subjects (IS): 7
- Educated Subjects (ES): 7

Above participants included females and left-handed subjects.

- Left-handed Subjects: 4

Educated: 1, Semi-literate: 2, Illiterate: 1

- Female Subjects: 4

Educated: 3, Semi-literate: 1, Illiterate: 0

The semi-literate and illiterate subjects included people with different professions like electrician, gardener, estate manager, farmer, driver and housewives.

4. TEST APPARATUS

Various PDA phones were observed such as HP IPAQ Pocket PC, HTC TYTN 4550, iPhone, etc. Out of these, HTC TYTN 4550 PDA phone with its stylus was chosen for testing. It was more suitable for testing due to smaller size of display (i.e. 2.8inch diagonally). HP iPAQ and iPhone displays are 3.5inch in size, which is quite large for one-handed thumb use. The specifications of HTC TYTN 4550 are given below-

Display: 2.8 inch, 240x320 QVGA TFT-LCD touchscreen

Operating System: Windows Mobile 6 Professional

Dimensions: 112 mm (L) x 59mm (W) x 19mm (T)

Our main intention of the study is to evaluate 'thumb' as the tool for interacting with the touchscreen and not to evaluate any specific model of PDA phone. Also, we do not intend to compare multi-touch or single touch aspects of touchscreens. Therefore, use of HTC TYTN 4550, as part of our experimentation, is only incidental. We chose it because of its smaller size, as that allows one to hold it in one hand and operate it using thumb.

5. DESIGN OF USABILITY TESTS

Agile approach to experimental testing is explored for identifying various usability issues centred on thumb interaction with smart phone. It meant incremental change in test design and conceptualization of new tests based on the findings. All experimental tests were designed using Adobe Flash CS 3 Professional and played through the PDA phone. An overview of experimental usability tests is presented below.

I. Initial experimental test was designed to evaluate the fine motor control of subjects and not the usability of thumb interaction. It involved multiple grids with varying sizes of buttons as shown in figure 1. The subjects had to click the buttons in serial order using stylus.

II. In continuation of above testing, when the button size became large enough to explore thumb interaction, the stylus was discontinued. We kept on increasing the sizes of buttons until it improved the performance results of thumb interaction.

III. A separate test was designed for testing the accuracy of thumb clicking and to identify the common patterns of

clicking. In this test, discrete clickable targets were presented in random locations on the screen of PDA phone.

After observing the results of initial three types of testing, the next set of usability tests were evolved.

IV. A piece of paper, matching the size of touchscreen, was stuck on the PDA phone. Thumb impressions were taken on this paper by applying ink to the tip of thumb. This test helped us in observing the contact area of thumb, its shape and angle, when it touched the screen. It also revealed the points on the tip of thumb and screen locations, where maximum force gets concentrated.

V. Buttons were arranged in a curved line, instead of usual straight-line arrangement of buttons. This test was designed to find the comfort and ease of use while clicking. Horizontally elongated shape of buttons was also tested in a similar way.

Total 12 usability tests were evolved out of which 10 of them are reported with details in this paper. Initial 6 tests involved gradual increase in the buttons size. The subjects had to perform the tests in a lab environment and in the sitting posture to minimize the complexities. Every subject was given enough time to practice before testing, until they felt comfortable with the medium. Details of each experimental test are elaborated in the sections ahead.

Please note that we focus on evaluating the usability of thumb as an interaction tool and not on the usability of interactive test applications.

6. EXPERIMENTS
(The graphs are generated using the examples from the actual performance by subjects.)

6.1 Test of Fine Motor Control Using Stylus

6.1.1 Test Description
As shown in figure 1. interactive applications containing varied sizes of grids with clickable buttons were prepared using Flash CS 3. Educated, semi-literate and illiterate subjects were asked to use the stylus for clicking the buttons provided in 28x28, 20x20 and 15x15 grids. Subjects were tested on each grid separately.
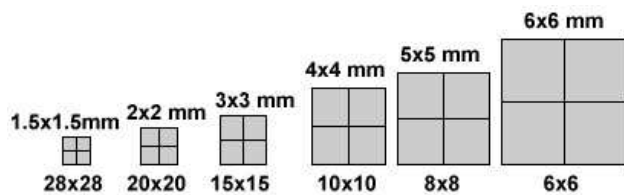

Figure 1. Different sizes of grids with clickable buttons

The behaviour of flash application was designed as under-

▪ Each test was conducted for exactly one-minute duration in sitting posture. The timer began ticking after the first click of the button in the grid. The time-out was indicated by displaying a blue circle on the screen.

▪ The subjects were given sufficient time to practice, until they felt confident, before starting to test. Each subject was tested separately.

▪ After clicking a button, it changed its color to red for feedback. Repeated click on the same button changed its color to green.
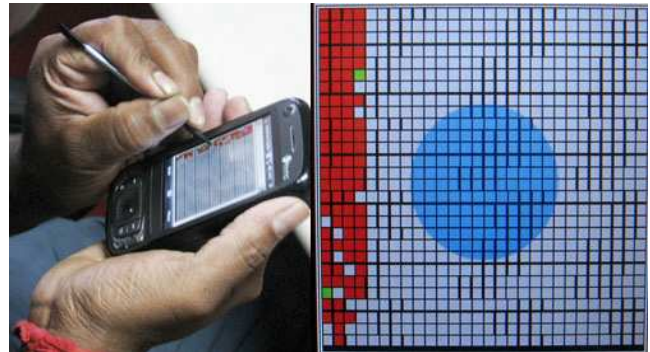
Figure 2. PDA with 28x28 Grid

▪ Subjects were asked to click the buttons in serial order with a comfortable pace, without the permission to return if they skipped some buttons. They were instructed to click with proper concentration on the task, without skipping any buttons.

▪ In this manner each subject was asked to click sequentially, in different directions and in different parts of the screen, as shown in table 2.

▪ After the time-out, photographs of final screens were taken for counting the total number of clicks and errors. Subjects were minutely observed when they performed the test.

Table 2. Test of sequential clicking in different directions	
Part of PDA screen	**Order of clicking**
Upper side of PDA screen	Left to Right (L-R)
	Right to Left (R-L)
Left side of PDA screen	Top to Bottom (T-B)
Right side of PDA screen	Top to Bottom (T-B)
Left side of PDA screen	Bottom to Top (B-T)
Right side of PDA screen	Bottom to Top (B-T)
Lower side of PDA screen	Left to Right (L-R)
	Right to Left (R-L)

6.1.2 Test Objectives
This test was designed with the following objectives.

1. Observe the difference between the fine motor control of educated, semi-literate and illiterate subjects in terms of the percentage of accurate clicks and errors (miss outs and repeated clicks).

2. Find out the subject's proficiency or lack of proficiency while clicking in different directions and different parts of PDA screen.

6.1.3 Observations
#### 6.1.3.1 General Observations
Initial tests on 28x28 grid with tiny buttons challenged the fine motor control and hand-eye coordination of subjects, as it

required greater concentration of eyes on the target and consistent incremental movement of the lower arm. Miss-outs and repeated clicks on same buttons represented the misjudgment of subjects. Subjects missed out some of the buttons while clicking, as their lower arm shifted a little more than necessary. Subjects became extra conscious while controlling the gradual movement of lower arm when they skipped a few buttons by mistake. This resulted in repeated clicks on same buttons, which is also counted as an error. We found higher frequency of such errors if subjects had poor fine motor control.

6.1.3.2 *Specific Observations*

Clicking in Left to Right order (Upper side of PDA screen)

- While clicking from left to right direction, we found that the error rate of semi-literate and illiterate subjects was much higher than educated subjects.

- The illiterate subjects became extremely conscious while holding the PDA and stylus in their shivering hands, and ended up clicking very slowly. This resulted in very less number of clicks.

- But the error rate of illiterate subjects became very high whenever they increased the speed of clicking as shown in graph 1. Total number of clicks within one minute by an illiterate subject are less by 64%, and that too with higher error rate, if compared with the educated subject.

- Educated subjects clicked more number of buttons speedily with few errors.
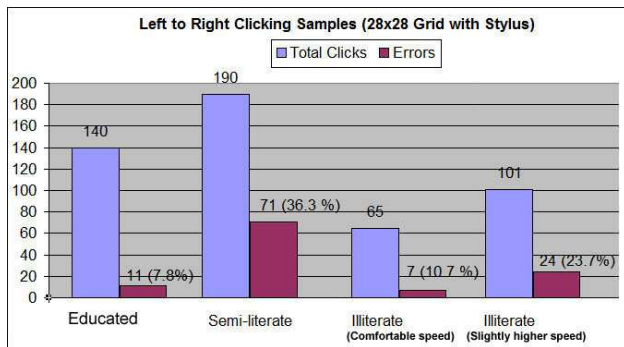
**Graph 1. Comparison of left to right clicking performance between the subjects**

Dexterity

- Left-handed subjects demonstrated greater proficiency while clicking the buttons in right to left order, if compared with their performance in left to right order as shown in graph 2.

- Right-handed subjects demonstrated greater proficiency while clicking the buttons in left to right order, if compared with their performance in right to left order.

- In case of right-handed illiterate subjects, their strength was not clearly visible in any of the tests. It showed lack of dexterity while using the stylus.

Clicking in Left to Right order (Lower side of PDA screen)

- All the subjects were not comfortable while clicking the buttons at the lower part of the PDA screen, as the hand does not get adequate support.

- On the contrary, while clicking in the upper portion of the screen, one was able to rest the hand on the edge of the PDA and hence their performance in this area was much better.
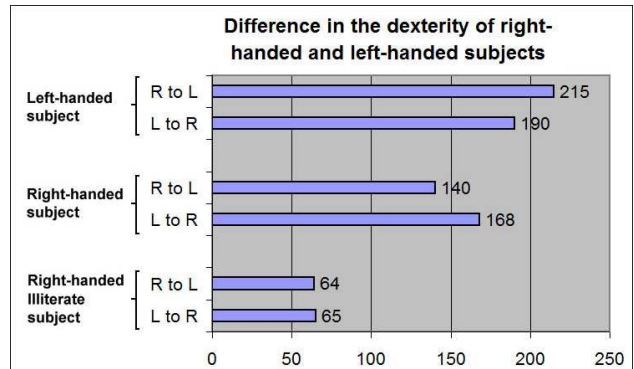
Graph 2. Difference in the dexterity of subjects (28x28 Grid)

Serial clicking in vertical order

- Top to bottom and bottom to top clicking sequences did not show any significant trend.

Slippery and thin stylus

- Figure 2. shows the photograph of how the illiterate subjects were holding the stylus in hand. The stylus was too thin and slippery for them to grip it properly between the fingers. It is also because they have seldom held the pen in hand for writing.

Fatigue

- We also found that subjects made more errors in the second half of the clicking tasks due fatigue.

Performance on 20x20 and 15x15 grids

- In the subsequent tests over slightly bigger size of grids such as 20x20 and 15x15, the performance of subjects improved marginally.

This test revealed that the stylus wasn't really a good option for the semi-literate and illiterate subjects due to their lack of fine motor control, which is developed out of writing practice. The tests hereafter were performed on 10x10, 8x8 and 6x6 grid using the thumb, as the button sizes were large enough to allow thumb interaction.

6.2 Interaction Using the Thumb

6.2.1 *Test Description*

In this test, all subjects were asked to click using their thumb instead of the stylus. 10x10 grid with buttons admeasuring 4mm by 4mm was used for this test. As shown in table 2., we followed the same pattern of clicking on PDA screen. The functional behavior of the application and rules were same as earlier.

6.2.2 *Test Objectives*

1. Find out whether semi-literate and illiterate subjects are able to perform better using their thumb than the stylus. Find out whether the fine motor control developed out of writing practice is helpful in thumb interaction.

2. Observe the error rate and compare it with subjects' performance using stylus.

3. Identify the areas on PDA screen where the subjects are more efficient and effective.

Figure 3. Thumb clicking by right handed (IS) and left handed (SLS) on 10x10 grid
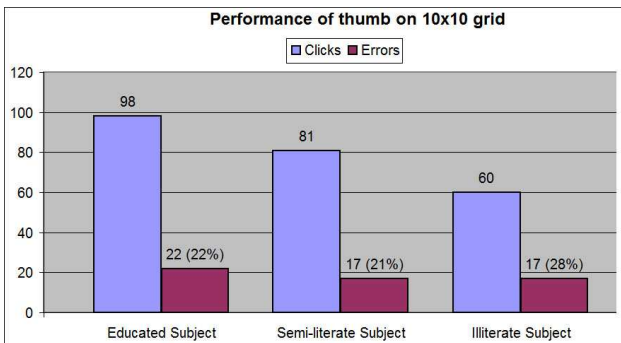
Graph 3. Performance using thumb on 10x10 grid

6.2.3 Observations

6.2.3.1 General Observations

All categories of subjects made greater number of errors while performing the test using their thumb. All subjects had to fumble due to ambiguous shape of the thumb. The performance of educated subjects also declined significantly in terms of accuracy if compared with their performance using the stylus as shown in graph 3. The interesting observation here is that there isn't contrasting difference in the performance of subjects, if we compare it with their performances using stylus, as shown in graph 1.

6.2.3.2 Specific Observations

Palm size and thumb length

- Subjects with smaller palm and shorter length of thumb faced difficulty in clicking the targets near the outside edge and upper left corner of the PDA screen. As they have to stretch the thumb which flattens the touch on the screen.

- Subjects with bigger palm and longer length of thumb faced difficulty in clicking the targets near the inside edge and lower right or left corner of PDA screen respectively (based on whether they are right-handed or left-handed). Refer figure 3.

Shape of thumb and fingernails

- Subjects with blunt shape of thumb made more errors.

- Subjects with properly shaped and long fingernails were more accurate. All semi-literate and illiterate subjects that participated in our test had slightly long and rounded fingernails. Refer figure 3.

- Educated subjects had short but properly shaped fingernails. They too were able to click properly.

- Small number of subjects with unkempt fingernails had to struggle a bit more for getting the proper touch on buttons.

Dexterity

- As shown in graph 4. all categories of subjects did not demonstrate dexterity in terms of their right-handedness or left-handedness when they used their thumb, if compared with graph 2. Their clicking performance in left to right and right to left sequences is almost similar. It means that the fine motor control developed out of writing practice wasn't helping the educated subjects, as much as it did while using the stylus.

Raised frame around the touchscreen

- The subjects faced difficulty while clicking the buttons near the raised frame around the touchscreen. Unlike iPhone, many other smart phones have a raised frame around the screen. It is not recommendable for one-handed thumb use.
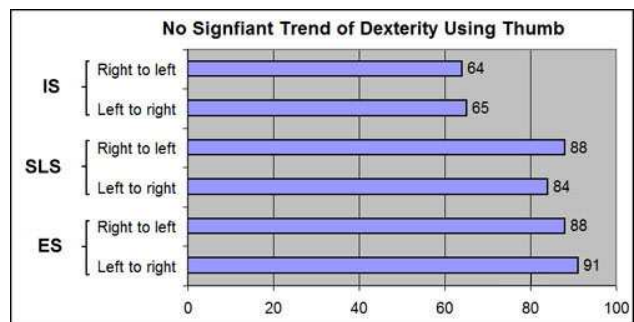
Graph 4. No significant trend of right or left-handed dexterity using thumb

Good performance in 8x8 and 6x6 grids

- Educated subjects and semi-literate subjects could easily click the buttons on 8x8 grid with minor errors. Here the diagonal measurement of button was 7mm.

- But the illiterate subjects found it difficult and made many mistakes while clicking on 8x8 grid. They were more comfortable with 6x6 grid and made almost no errors. Here the diagonal measurement of the button was 9mm.

- Looking at the consistent trend of improvement throughout the testing, we believe that with more practice of thumb interaction, the semi-literate and illiterate users will be able do as good as educated users.

The usability tests hereafter are focused on understanding what makes the thumb an ambiguous and imprecise tool for interaction.

6.3 Accuracy of Clicking Using Thumb

6.3.1 Test Description

This test involved a grid of 15x15 clickable buttons. However, only the 5x5 grid is made visible to the subjects. Each box in the grid actually consists of 3X3 (total 9) clickable buttons, which are not outlined. One tends to perceive the group of 9 buttons as the single target to be clicked using thumb.

In this test, each box appeared on screen after 3 seconds and continued to remain there till end. Every time, the new box

appeared in a different location of screen. In this manner, the subjects were asked to click all 25 boxes. The subjects had to perform this experiment three times.
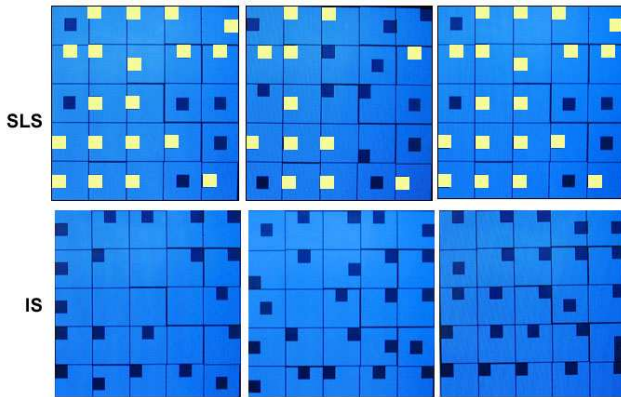

Figure 4. Pattern of thumb clicking

As each box is consisting of 9 clickable buttons. While clicking using the thumb, the area where maximum force was concentrated, the button underneath got clicked and changed its color to dark blue.

6.3.2 Test Objectives
- Evaluate the preciseness of clicking using the thumb. Understand the reasons behind why the subject was unable click exactly in the center of target in certain locations.
- Find out whether there are any common patterns in the clicked points and locations on PDA screen.

6.3.3 Observations
Pattern
- We found that each person has a common pattern of clicking using the thumb, which consistently repeated.
- As shown in figure 4. (first row of three photographs) the pattern of clicking is indicated by highlighting the boxes in yellow color. Each time the subject has tried to click the box in its center but in certain locations (s)he has ended up clicking either on left side of the box or at the upper portion of the box. The second row of three photographs in figure 4., show the performance by an illiterate subject.

Common miss outs
- Many times the subjects failed to click in the center of target although their intention was always to click in the center. It shows the unpredictability of thumb as a pointing device.

6.4 Thumb Impressions on PDA Screen

6.4.1 Test Description
The earlier test showed that there is consistent pattern in the way thumb touched the PDA screen in different locations. Therefore, we decided to experiment with another test, in which ink was applied on the thumbs of subjects. A piece of paper exactly matching the screen size was stuck on the PDA. The subjects were asked to touch different locations on this paper, while holding the device in one hand. Thumb impressions that appeared on paper were used for analysis.

6.4.2 Test Objectives
- Observe the contact area of thumb in different locations of screen; and its changing shape and angle based on the length of thumb.
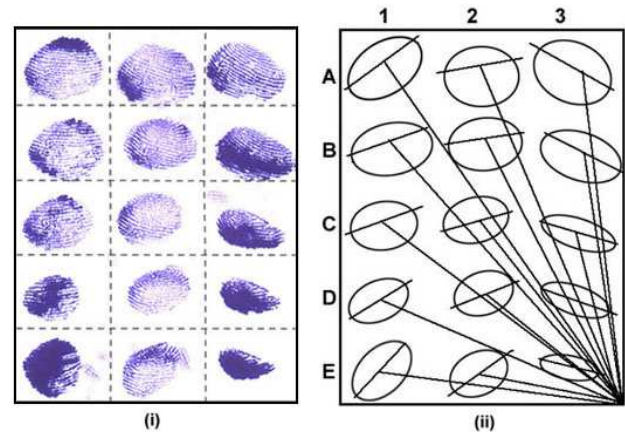

Figure 5. (i) Thumb impressions reveal the contact area in different locations of screen (ii) Asymmetric movement of thumb from lower right corner

6.4.3 Observations
Flat and pointed finger touch
- During one-handed thumb use on PDA screen, certain locations are conveniently within the reach of the thumb; and for certain locations the thumb has to be stretched.
- As shown in figure 5(i), as per thumb impressions, the contact area of the thumb towards the lower right corner of the screen has become elongated and narrow. Whereas, the contact area of thumb towards the top left corner and middle of the screen, is quite large and oval shaped.

Stretching and shrinking for length adjustment
- Figure 5(ii). shows that the thumb needs to be stretched to reach the locations A1, A2, A3 and B1 for those with shorter length of thumb. Such stretching can be stressful.
- Whereas, the thumb has be folded and shrunk in its length to reach the locations E3, E2, D3, D2, D1 and C1. Such shrinking can be stressful for those having long length of thumb.
- Locations like E1, D1, C1, C2, B2, B3 are comfortably reachable for most subjects.

Thumb approach from a corner of PDA screen
- It is most important to note that the thumb approaches the PDA screen from lower-right corner (if right-handed) and lower-left corner (if left-handed). This also results in asymmetric movements.

Asymmetric movements
- Thumb approach is from a corner of the screen, which causes the thumb to move in an asymmetric way. As shown in figure 5(ii), While moving from target A1 to A3 one has to gradually shorten the length of thumb.

- On the contrary, the length of stylus remains constant while approaching any of the targets on the screen. This is the major difference in thumb and stylus based interaction.

Circular motion of thumb

- The thumb has a natural circular movement centered around its joint that connects with the palm. But we forcibly move our thumb linearly in horizontal direction on the screen. It can stress the thumb joints and cause Repetitive Stress Injuries (RSI) in the long run.
- Also, the contact area of thumb is oval shaped.
- As shown in figure 7., we can evidently notice circular, wiper like movement of the thumb, which is most natural.
- But contrary to this, the user interface, menus, and buttons are organized in linear order. It forces the thumb to be moved in straight line, against its natural circular motion.

6.5 Physical Study of Thumb

Some common observations are already mentioned in point 6.2.

We measured the thumbs of subjects based on the length of its phalanges and the circumference of the distal phalange (tip of the finger) as shown in figure 6(i). Sample measurements are given in table 3.
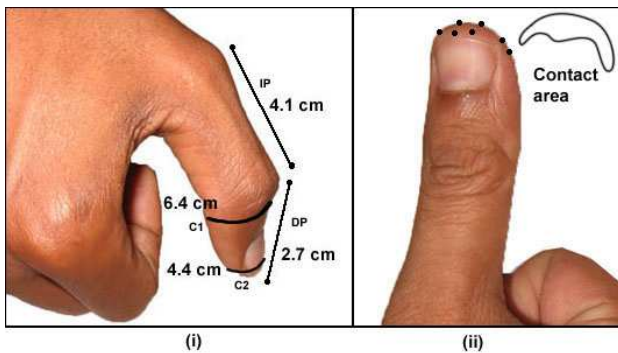
Figure 6. (i) Measurements of thumb (ii) Different contact points of thumb on the touch screen

Blunt distal phalange

Even within the small sample of thumbs measured by us we found some exceptions where the circumference of the distal phalange (C2) was slightly larger by couple of millimeters than the other comparable thumbs. Refer table 3. We found these subjects struggling a bit harder than others, while clicking the targets.

Pointed Distal Phalange

We also came across many subjects who had pointed thumbs, with significantly less circumference of the distal phalange.

Asymmetrically distributed and shifting points of contact

As shown in figure 6(ii) the tip of the thumb touches the screen at multiple points. Thumb is not a single point device like the stylus. Mostly the points of contact on the thumb move along the horizontal axis and they are distributed more on the right half of the tip for right-handed users.

Impact of occupation

Softness of the palm skin and the shape of fingernails also depend on the occupation of subjects e.g. farmers and factory workers

have rough hands. All educated subjects had softer hands if compared with the semi-literate and illiterate subjects.

Short-forms used in table 3.
- Length of Intermediate Phalange (IP), Length of Distal Phalange (DP)

Measurements of distal phalange
- Circumference 1 (C1)
- Circumference 2 (C2)

Table 3. Sample measurements of thumbs in centimeters			
Female Thumbs			
IP	**DP**	**C1**	**C2**
3.7	2.9	5.1	3.7
3.5	2.7	5.2	3.4
<u>3.7</u>	<u>2.8</u>	<u>5.4</u>	<u>**3.4**</u>
<u>3.7</u>	<u>2.9</u>	<u>5.6</u>	<u>**3.9**</u>
Male Thumbs			
IP	**DP**	**C1**	**C2**
4.1	3.4	6.4	4.4
<u>3.9</u>	<u>3.5</u>	<u>6.3</u>	<u>**4.7**</u>
<u>3.8</u>	<u>3.4</u>	<u>5.7</u>	<u>**4.1**</u>
<u>3.8</u>	<u>3.4</u>	<u>6.1</u>	3.9

7. DESIGN RECOMMENDAIONS

7.1 Challenges of Thumb Interaction

Factors of thumb that impact the interaction

Length of thumb, circumference of the tip of thumb, length and shape of fingernail, size of palm, softness of skin, the location of approach to the PDA screen (lower-left corner or lower-right corner), its contact point and contact area on the PDA screen are the variable factors that impact the quality of interaction.

Why is thumb clicking ambiguous?

The shifting contact points and varying dimensions of contact area in terms of its shape, size and angle make the thumb an ambiguous and imprecise tool for interacting with smart phones. The contact points shift and the contact area changes mainly because the thumb approaches from the corner of PDA screen.

7.2 Improving Thumb Precision and Ease

The observations and insights of experimental tests have helped us come up with following recommendations for improving the effectiveness of thumb interaction.

Elongated shape of buttons

It is noticeable that the multiple contact points on the tip of thumb, are mostly distributed in horizontal order, more towards the right half of the thumb for right-handed subjects. Therefore it is logical to provide horizontally elongated shape of targets for better results.

As shown in figure 7, the curved arrangement of buttons was most comfortable and easy to all subjects. They unanimously felt that

the circular motion of thumb was more natural than the straight-line movements on the screen.

Capturing the size of contact area to decide the target size

Universal and fixed target size may become too small or too large for many users. Very large size of target may be constraining for complex applications. Therefore, it will be ideal to intelligently capture the size of contact area of thumb (location wise) for every user and then adjust the target sizes. This type of personalization may be essential at least for the productivity tools.

Selectable design preference

Selectable user interface layouts and the suitable inclination of screen should be provided for left-handed and right-handed users. They are very important in the context of one-handed thumb use, as the reachable locations on touchscreen are different for both types of users.

Inclined screen for center alignment with thumb

As shown in figure 8, if we consider the angular position of the thumb, struggling to use it over a perpendicular screen can be definitely difficult, unnatural and stressful. The inclined screen seems like an ergonomic proposition for one-handed thumb use. It can bring certain amount of symmetry to the movement of the thumb. If the screen is positioned at an angle, in proper alignment with the thumb, the curved arrangement of buttons is not necessary.

Figure 7. Curved arrangement of buttons
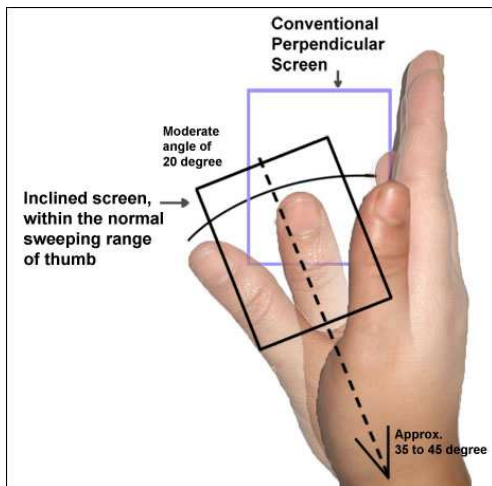
Figure 8. Natural angular position of thumb and the proposed screen alignment

Size of screen and the device

As shown in figure 8., the size of screen should ensure reachability of thumb in all locations. The raised frame around the screen should be avoided. The width of the device should be such that it allows one to properly hold it in one hand.

Improved preciseness and thumb control for the users

If the thumb can approach the PDA screen from its centre of the bottom edge as shown in figure 8., the contact areas and contact points will get symmetrically distributed. This can improve the predictability and precision of thumb interaction.

Reduction in stress injuries

If the screen is inclined to align with the thumb, then movements will become natural, symmetric, easy and ergonomically suitable. It will reduce the stretching and shrinking of the thumb joints at odd angles.

8. DISCUSSION

The test of fine motor control using the stylus evidently shows that it gives greater advantage to educated subjects, if compared with the semi-literate and illiterate subjects. It is mainly because educated subjects have good writing practice. The fine motor control developed through writing practice is marginally helpful while using the thumb for interaction. The other important observation is that there isn't contrasting difference between the performances of educated, semi-literate and illiterate subjects, when they used their thumb for clicking. This is primarily because it is natural for all human beings to use their thumb quite proficiently irrespective of their level of education.

We found that each person has a unique pattern of clicking using the thumb, which consistently repeats. This happens as the thumb approaches the screen from a certain location and it has a fixed length and shape. One tends to misjudge the target because of the shape of thumb and its angular approach to the screen. Constant activity of stretching and shrinking of the thumb and angle adjustment to get the proper touch point on the screen, can put major stress on the thumb joints. Thumb kinematics [16] is really put to test when we use our thumb on the touchscreen of smart phone.

Although we found that semi-literate and illiterate subjects could easily click the targets in 8x8 and 6x6 grid respectively, it is difficult identify one size of buttons, which will suit all. The suitable button size i.e. 9mm found out by us, is smaller than what is recommended by Karlson [8] and Parhi et al. [15]. The anthropometric data of every country shows different anatomical proportions and average heights [24]. Universal size of button can be too large or too small for users. Therefore, it would be ideal to intelligently capture the contact area of thumb in different locations of screen and adjust the target sizes accordingly.

The multiple contact points on the tip of thumb are spread horizontally. They are more on the right half of the thumb for right-handed users. It shows asymmetry and imbalance due to angular approach of thumb. It indicates that horizontally elongated shape of buttons can improve the performance. To remove the asymmetry and imbalance in the position and circular motion of thumb, it is suggested to incline the screen by 20 degree, so that thumb can approach the touchscreen from its center of the bottom edge. It will help in reducing the stress and simplifying the kinematic adjustments [17] required for clicking various targets.

Another interesting observation is that while traveling in a vehicle, coordination between two hands holding stylus and PDA separately goes haywire due to shakiness. In such situation, one-handed thumb interaction proves more effective, as the hand that holds PDA and the thumb that interacts with it, both move together simultaneously. They are one entity.

9. CONCLUSION

One-handed use of thumb as the tool for interacting with smart phone is recommendable for semi-literate and illiterate users. Their lack of fine motor control (skills developed out of writing practice) does not adversely impact their use of thumb on touchscreen. It is essential to incorporate the design recommendations emerged out of the usability study to further improve overall effectiveness of one-handed thumb use. With these enhancements, it will be possible to tap the potential market of mobile-based occupational productivity tools.

10. ACKNOWLEDGMENTS

11. REFERENCES

[1] Balakrishnan, V., Yeow, P. A Study of the Effect of Thumb Sizes on Mobile Phone Texting Satisfaction, *Journal of Usability Studies*, Vol.3., Issue 3, May (2008), Pages 118-128.

[2] Brewster, S., Chohan, F., and Brown, L. Tactile Feedback for Mobile Interactions, *Proc. CHI* (2007), Pages 159-162.

[3] China and India Lead the Worldwide Telecom Carrier Capex and Revenue Growth, Celluar-news, July (2008) http://www.cellular-news.com/story/32251.php

[4] Cavalli-Sforza, L., L. Origin and Differentiation of Human Races, *Proceedings of the Royal Anthropological Institute of Great Britain and Ireland*, No. 1972 (1972), Pages 15-25.

[5] Fitts, P., M. The information capacity of the human motor system in controlling amplitude of movement. *Journal of Experimental Psychology*, *47* (1954), 381-391.

[6] Gokturk, M., Sibert, J. An Analysis of Index Finger as a Pointing Device, *Proc. CHI*, (1999), Pages 286-287.

[7] Hoggan, E., Brewster, S., Johnston, J. Investigating the effectiveness of tactile feedback for mobile touchscreens, *Proc. Conference on Human Factors in Computing Systems*, (2008), Pages 1573-1582.

[8] Karlson, A., K., Interface Design for Single-handed Use of Small Devices, USIT 2006 Adjunct Proceedings: Doctoral Symposium, (2006), Pages 29-30.

[9] Karlson, A., K., Bederson, B., B. and SanGiovanni, J. AppLens and LaunchTile: two designs for one-handed thumb use on small devices. *Proc. CHI 20*05, ACM Press (2005), 201-210.

[10] Katre D., Cross-cultural Usability of Bilingual Mobile Phones, *Proc. Indo-Danish Symposium on Cultural HCI*, India, (2006).

[11] Kurniwiawan, S. Mobile Phone Design for Older Persons, ACM Interactions, Volume XIV 4., July + August (2007), Pages 24-25.

[12] India And Pakistan Gearing Up For Mobile Data Revenues (July 2008) http://trendsniff.com/2008/07/31/india-and-pakistan-gearing-up-for-mobile-data-revenues/

[13] Monthly Economic Analysis, Fortune 2008, *A Report published by Ministry of External Affairs, Government of India*, May (2008), Pages 5-6.

[14] Mizobuchi, S. Mori, K., Ren, X. and Yasumura, M. An empirical study of the minimum required size and the number of targets for pen on the small display. *Proc. MobileHCI* 2002, Springer-Verlag (2002), Pages 184-194.

[15] Parhi, P., Karlson, A., Benderson, B., Target Size Study for One-handed Thumb Use on Small Touchscreen Devices, *Proc. Conference on MobileHCI*, (2005), Pages 203-210.

[16] Sandova G, Falling revenue for cell phone maker (iSuppli Report), CNET News.Com, January (2006) http://news.cnet.com/Falling-revenue-for-cell-phone-makers/2100-1039_3-6026127.html

[17] Santos, V. J. and Valero-Cuevas, F. J. Thumb kinematics with non-orthogonal and nonintersecting axes of rotation may be necessary to predict realistic isometric thumbtip forces in multiple directions. Proceedings of the 28th Annual Meeting of the American Society of Biomechanics, Portland, OR, Sept. 2004.

[18] Sears, A., Revis, D., Swastiki, J., Crittenden, R. and Shneiderman, B. Investigating Touchscreen Typing: The effect of Keyboard Size on Typing Speed. *Behavior and Information Technology 12, 1* (1993), Pages 17-22.

[19] Siek, K., Rogers, Y., Connelly, K. Fat Finger Worries: How Older and Younger Users Physically Interact with PDAs, *Interact* (2005)

[20] Singh, A., P., Singh, S., P., Impact of Habitual Physical Activity on the Human Body Proportions: A Comparative Study of Some Traditional Groups, *Journal of Human Ecology*, (2007) 22(3): Pages 271-275.

[21] The World Factbook, India https://www.cia.gov/library/publications/the-world-factbook/print/in.html

[22] Worthen, B., Are Mobile Workers at Risk of BlackBerry Thumb?, *Business Technology, Wall Street Journal*, June, (2008).

[23] Wu F. G., Luo S. Performance study on touch-pens in three screen tasks, *Applied Ergonomics*, Elsevier (2006), Pages 149-158.

[24] Human Height, Wikipedia http://en.wikipedia.org/wiki/Human_height

[25] Yunferi, X., Ergonomic Experiment for Thumb Keyboard Design, *IEEE CAIDCD (2006)*, Pages 1-5.

Evaluation of Chinese Finger Writing Recognition on Touch-Sensitive Keypad of Mobile Phones

Huanglingzi Liu
Nokia Research Center
No.5 Dong Huan Zhong Lu, BDA Beijing
linda.2.liu@nokia.com

Ying Liu
Nokia Research Center
No.5 Dong Huan Zhong Lu, BDA Beijing
ying.y.liu@nokia.com

ABSTRACT
Finger handwriting based on the standard-looking keypad was a unique text input technology. A usability test was carried out on it in this study. Some unique usability problems relative to finger input approach and the confusable activity of finger press and finger touch are analyzed. Based on the analysis of the evaluation result, the part of concept design provides a new solution to improve finger input usability and user satisfaction.

Categories and Subject Descriptors
H.5.2 [**Information Systems**]: Information Interfaces and Presentation – *Evaluation, Input devices and strategies*

General Terms
Measurement, Design, Human Factors.

Keywords
Usability, text entry, finger, handwriting recognition, Chinese

1. INTRODUCTION
Touch technologies are applied in many handheld products. Touch screen, as a key touch-sensitive device, received much attention from both practitioners and researchers. Many new interaction solutions were provided based on the technology, such as multi-touch from iPhone or iPod. However, touch screens are expensive and perceived as a device that is lack of precision [4]. In many cases, a touch pad is a good substitution of a touch screen for some simple tasks. For example, Nokia 6108 enabled Chinese consumers to enter Chinese characters with a touch pad. Ethan Cheng et al also developed a Chinese text entry system and enables Chinese users to enter characters by writing phonetic symbols with a capacitive touch-sensitive keypad [1].

The disadvantages of a touch pad are also straightforward. Touch pad itself cannot present either visual or tactile feedbacks, which means that it cannot be the single input and output device in a product as a touch screen can be. Some local phone products in China provide a capacitive touch-sensitive keypad with the 12 keys so that users can enter Chinese characters by writing

characters with their fingers on the touch-sensitive keypad. The keys of the keypad are thus very flat to make sure users can write Chinese characters smoothly. The combination brings very unique experiences to end users. No published results are available to report usability of it.

Moreover, handwriting recognition is a preferred way of entering Chinese characters by many local end users. It is a natural solution for entering Chinese. Not all Chinese users are familiar with pinyin or stroke methods based on the 12-key keypad. For them, handwriting recognition is a good substitution.

The importance of Chinese handwriting recognition and the unique way of finger writing recognition on touch-sensitive keypad drive us to conduct the study. The objectives of the study were:

- to compare usability of the finger handwriting recognition solution on keypad with a similar handwriting recognition solution with pen and a touch pad;

- to check other usability problems of the solution and propose new concepts to improve it

2. METHOD

2.1 Test Devices
We chose two mobile phones to study the finger handwriting recognition solution: Motorola A732 and Samsung W399. Nokia 6108 was chosen as the comparative device with the handwriting recognition solution of pen and touch pad.

Figure 1 shows pictures of the three devices evaluated in the study. Motorola A732 is a slide phone. Beside finger handwriting recognition, it also supports some basic gesture commands for editing and mode switch among languages. Moreover, pressing the "#" key also enables users to change input mode among languages. The size of finger input area is around 34 mm's wide and 29mm's height. Samsung W399 is a fold phone. It doesn't support gesture input. Users can press the "*" key to change input mode. The size of finger input area is around 40mm's wide and 36mm's height.

Nokia 6108 is a good comparative device for finger handwriting recognition solutions. It support pen based handwriting recognition on a touch-sensitive pad. The other difference is that in Nokia 6108, the touch pad is separate from the keypad. In addition to the writing area in the touch pad, some functional keys including "ok", "clear" and mode change keys are also labeled and presented. Some simple gestures are also supported by the Nokia 6108. For example, drawing a line from right to left can

delete a character. The size of the input area for Nokia 6108 is around 19mm (width)*20 mm (height).

a. Motorola A732   b. Samsung W399   c. Nokia 6108

Fig. 1. Phones supporting handwriting recognition with touch-sensitive keypad, which are evaluated in the study

2.2 Participants

8 users, half male and half female, volunteered to attend the usability study. Their ages range from 18 to 37 years. All are right handed. 6 of the participants are familiar with keypad based text input; 2 are novice users of keypad based text input. 3 participants are familiar with pen text entry solutions with touch screens. 1 participant is the experienced user of Nokia 6108 and the other 4 are novice with handwriting recognition. The average width of the index fingers of the 8 participants was 13mm (s.d. 1.18).

2.3 Task design and Procedure

At first participants were introduced the test objectives and procedure. Then the participants were asked to fill out a pre-test questionnaire, which is to collect past experience on mobile phone and text entry solutions.

A within-subjects design was applied in the study. All participants need to complete tasks with all the three devices. To avoid the carryover effect, we designed three similar text entry tasks. With each task, a participant needed to enter Chinese characters, numbers and English letters. The order of test devices and the tasks was balanced among the 8 participants.

Fig. 3. Usability test procedure

Before the real test session, all participants were instructed to play approximately five minutes with the device and enter simple sentences. In the real test session, participants were asked to 'think-aloud' if there was any problem. All the input processes were recorded with a video camera. After completing the text entry task with one device, the participants filled out a post-task

questionnaire. After the participants finished all the three tasks, they filled out a post-test questionnaire regarding preferences and satisfaction degrees. The whole test session took approximately one and a half hours.

3. RESULTS

3.1 Finger Handwriting vs. Pen Handwriting

Figure 4 and table 2 show the results on text entry rates. ANOVA test indicated that the three devices are significantly different on text entry rates. Further T test found that the pen handwriting recognition solution was significantly quicker than the other two finger writing recognition solutions (Nokia 6108 vs. Moto A732: t=2.542, p<0.05; Nokia 6108 vs. Samsung W399: t = 4.096, p<0.01). There was no significant difference between the two finger handwriting recognition solutions (t=1.424, ns).

Table 1. Chinese character input speed with three different models

	Text entry rates			F	P
	Moto a732	SamSung w399	Nokia 6108		
Mean	6.3	6.6	11.2	5.630	0.011
SD	3.28	2.03	3.16		
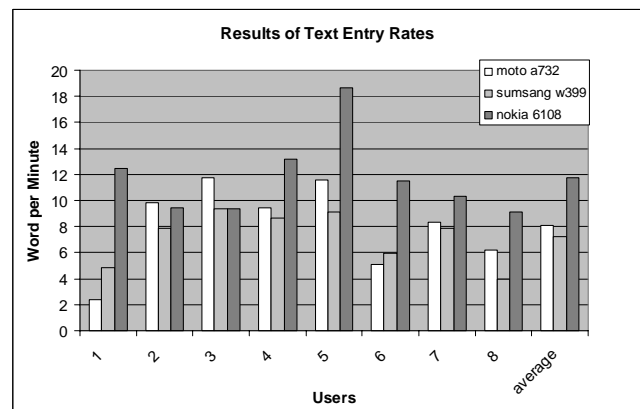
Fig. 4. Chinese handwriting speed with 3 test devices

Participants evaluated the three solutions after they finished the text input tasks. Figure 5 shows how 8 participants averagely scored the easy-to-use, satisfaction and efficiency level in five-point Likert scale. Based on the average scores, pen handwriting is regarded more satisfied, efficient and easier to use compared to finger handwriting. But ANOVA test indicated that the three devices are not significantly different on the three subjective evaluation items. The F values on the subjective easy-to-use, satisfaction, efficiency level are 1.90 (p=0.17), 2.21(p=0.13) and 1.02 (p=0.38) respectively. A main good point with finger handwriting we found in the comparison is that although all the participants in this study are novices in finger handwriting, they didn't think finger handwriting is difficult. The subjective easy-to-use level of finger input solution on both Moto and Samsung is higher than 3.
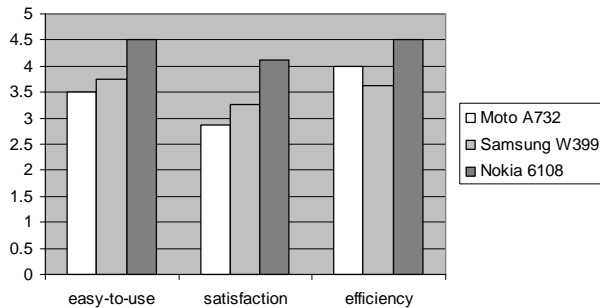
Fig. 5. Comparison of the subjective evaluation

After completing the test with each device model, participants were required to fill out a post-task questionnaire to collect their subjective evaluations on some UI design factors. In this study, we mainly explored five factors: writing power requirement, finger input size, synchrony of stroke trace, size of stroke trace and time out. The effect of the above five factors on the subjective evaluation on the ease level of text entry was also checked by analyzing the correlation between the two categories of variables. Only the correlation between the finger input area size and the ease level of text input was significant and the Kendall coefficient is 0.51 (p=0.001).

3.2 Usability Problems with Touch Sensitive Keypad Based Finger Handwriting

3.2.1 Small Finger Handwriting Input Area
One general error that we observed with the finger handwriting recognition solutions was that the participants often wrote outside of the input area. Since the recognition would start once a timeout reached a pre-set amount, system started to recognize the unfinished characters.

The errors happened more often with complex Chinese characters, which usually are composed of many strokes and a few components. End users have big difficulty in writing a whole character with relatively cumbersome finger tips. Users also wrote strokes overlapping with each other, which surely decreases the recognition rate.

The human finger as a pointing device has very low 'resolution'. Many studies explored proper sizes for finger interactions [4,5,6]. Parhi et al. studied the required object sizes for discrete and serial pointing tasks with thumb when a mobile device was held by one hand [6]. They concluded that the sizes of 9.2 mm and 9.6 mm were fairly large enough for discrete and serial pointing tasks. There are also some findings showing that it is difficult to point targets that are smaller than the finger width [5]. In our study, users also subjectively complained that the sizes of the input areas for fingers are not big enough.

3.2.2 Separate Touchpad and Display Screen
The keypad cannot provide visual feedbacks, which means users can not see what they wrote at exactly the same place where they wrote. Participants usually get feedbacks by the handwriting trace shown on the display. In such cases, it is impossible for the participants to see the accurate spatial distance between the going-to-be-written stoke and the existing stokes. The inter-stroke

distance, stroke connection and spatial positions between strokes directly influence Chinese handwriting recognition result [7]. Different stroke length may result in two different characters. For example, the only difference between '由'and'申' is the different length of the upright stroke of '|'. Figure 6 showed a true example that we observed in the study.

Fig. 6. The user has to rewrite the Chinese characters carefully since some stokes he input earlier were deformed

3.2.3 Combined Touchpad and Keypad
The third general error is related to mode errors. The combination of touch pad and keypad inevitably led to mode errors. For example, a press of a key may activate finger handwriting recognition if users happened to move their fingers a bit before clicking the key. This is especially true when a finger writing system allows users to select the right character by clicking the number keys on the keypad.

Besides the combination of the keypad and touch pad, the various language modes and the switch methods by pressing the "*" key made the condition even worse. First, users always forgot to switch language mode; Second, when users switch mode by clicking the "*" key, they sometime accidentally activate the finger writing recognition and led to errors. A lot of actions are needed to correct such errors.

Comparing the two finger handwriting solutions, the severity level of mode errors with Moto A732 seems minor than with Samsung W399. With Moto A732, under Chinese handwriting input mode, users are not supposed to press numerical key to select candidate characters. Pressing numerical key at that context may trigger keystroke Chinese input, while long press of some numerical key is to select the candidate character. According to our observations, nearly all participants gave up using numerical key to select candidate character but use navi-key instead. On the other hand, with Samsung W399, under Chinese handwriting input mode, users can press numerical key to select candidate character and input numbers without switching text input mode. Although users may generate a stroke ink during the process, most of them continued this operation instead of giving up; they thought it is more convenient and intuitive. It seemed that mode error rate could be decreased but at the cost of use flexibility or convenience.

4. DESIGN AND DISCUSSIONS
As to the last two general usability problems, there are already some design guidelines. Jun Rekimoto et al proposed PreSense which is a keypad enhanced by proximity sensors based on capacitive sensing [3]. With PreSense, users can be informed 'what will occur as a result of some key press.' This 'previewable user interface' can help to solve the usability problem mainly caused by the separation of handwriting input area and display screen. Users can regulate their handwritings if they know the

position of the new stroke is not attuned to the whole structure before they actually write down the stroke. Recognition rate can be improved by this way. Jef Raskin proposed "quasimodes", which are modes that are kept in place only through some constant action on the part of the user [2]. Modifier keys on the keyboard, such as the Shift key, the Alt key and the Control key, are all examples of a quasimodal interface. Same action will always produce the same perceived result. According to the design guideline, a special 'quasimode key' needs to be combined with the target alphanumerical key to fulfill some key press triggered functions. With this solution, users should not have to remember the current text input mode when inputting English letters or numbers.

However, as to the first usability problems, there is no special solution yet since required handwriting input size for cumbersome fingers is not compatible with device portability. We then proposed a new solution which makes use of one important feature of Chinese characters. Chinese characters are usually composed of different blocks.

Usually there are 3 ways to divide a Chinese character:

- one into left and right blocks. Examples are: '树','翻'

- one into top and bottom blocks. Examples are: '翼','囊'

- one into outside and inside blocks. Examples are:'国',' 癌'

The new solution enables end users to divide a complex character into different blocks, write them individually and then recognize them as a whole.
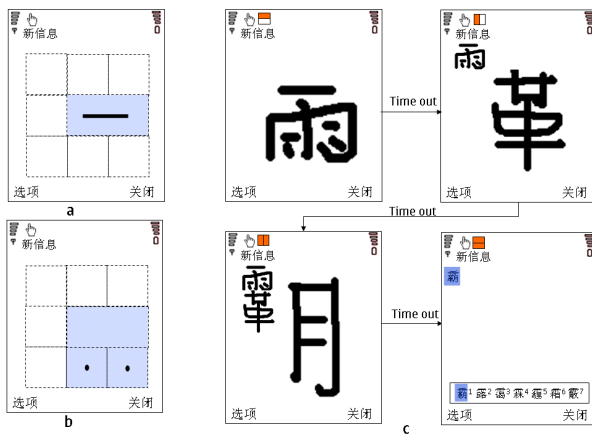
Fig. 7. The user writes the three blocks '雨', '革' and '月' to input the Chinese character '霸'

One example for the implementation could be as follows (see figure 7). End user firstly defines the structure of a complex Chinese character with a nine-square grid, which means how the user takes the character apart. Several adjacent squares can be merged into one square by dragging a line through these squares. Single square is selected by tapping. The selected square means one to-be-combined writing block. For example, if end user wants to write 霸 by '雨', '革' and '月', he/she firstly drags '一' like stroke in the nine-square grid (fig.7.a), then tap separately two squares just under the merged one (fig.7.b). Then end users can write the components one by one (fig.7.c). Following writings can

be combined according to the structure defined before and recognized as a single character at one time instead of being recognized separately.

5. CONCLUSION

A usability evaluation of finger handwriting with touch sensitive keypad was conducted. We tested and compared three devices, which were Moto A732, Samsung W399 and Nokia 6108. Comparison on text entry speed indicates that pen handwriting is more efficient than finger handwriting. But the subjective evaluation results show there are no significant differences on the easy-to-use, satisfaction and efficiency level among the three solutions. Although all the 8 participants in this study are novices in finger handwriting, they didn't think finger handwriting task is difficult.

The small handwriting area available on mobile phone keypad and the comparatively bigger size of finger tip prohibit smooth input and cause time out errors. Confusion about using finger or key on the same panel and separation of input and out area also lead to specific errors of touchable keypad based finger handwriting.

A new design is proposed in order to overcome the limits of finger writing in a small touchpad. Quick expert evaluation shows that there may be new potential usability problems with the new solution. Further prototype development and evaluation needs to be done.

6. REFERENCES

[1] Ethan Cheng, Renaud Amar, Nada Matic, Chinese Phonetic Handwritten Text Input for Mobile Phones,UIST'05

[2] John Rushby. Using Model Checking to Help Discover Mode Confusions and Other Automation Surprises. Proceedings of the Workshop on Human Error, Safety, and System Development (HESSD) 1999.

[3] Jun Rekimoto, Takaaki Ishizawa, Carsten Schwesig, Haruo Oba, PreSense: Interaction Techniques for Finger Sensing Input Devices, Proceedings of the 16th annual ACM symposium on User interface software and technology

[4] Mizobuchi, S., Mori, K., Ren, X. and Yasumura, M. An empirical study of the minimum required size and the number of targets for pen on the small display. Proc. MobileHCI 2002, 184-194.

[5] Pär-Anders Albinsson, Shumin Zhai, High Precision Touch Screen Interaction,CHI2003

[6] Parhi, P., Karlson, A. K., Bederson, B. B. Target size study or one-handed thumb use on small touchscreen devices, proceedings of the 8th conference on Humancomputer interaction with mobile devices and services, (2006), 203-210.

[7] Wing-nin Leung, Kam-shun Cheng, A stroke-order free Chinese Handwriting Input System Based on Relative Stroke Positions and Back-Propagation Networks, Proceedings of the 1996 ACM symposium on Applied Computing

[8] Ying Liu, Kari-Jouko Räihä, RotaTxt: Chinese pinyin input with a rotator, MobileHCI2008, 225-233

A Study of Usability Problem Finding in Cross-Cultural Thinking Aloud Usability Tests

Qingxin Shi
Copenhagen Business School
Department of Informatics

Howitzvej 60, DK-2000 Frederiksberg, Denmark

+45 72421536

Qs.inf@cbs.dk

ABSTRACT

Considering the cross-cultural cost-benefit analyses, sometimes it is more cost effective and more efficient to use a foreign evaluator instead of a local one. This study investigated how local and foreign evaluators found usability problems and rated the severity of the problems in thinking aloud usability tests in Denmark and China. The results showed that the number of usability problems found by local and foreign evaluators was similar. But the way of finding the problems and the description of some specific problems were different. A unique problem which was found only by foreign evaluators, not local ones, was also discussed. The analysis of the severity rating showed that Chinese evaluators preferred to rate the problems as "important" more than "minor" or "critical", whereas Danish evaluators did not have a clear preference of rating the problems.

Categories and Subject Descriptors

H5.2 [**Information interfaces and presentation (e.g., HCI)**]: User Interfaces--*Evaluation/methodology.*

General Terms

Human Factors

Keywords

Thinking Aloud Usability Testing, Culture, Usability Problem.

1. INTRODUCTION

Culture plays a more and more important role in the global market nowadays. There have been many studies about culture's influence on products or interfaces [1-4]. An increasing number of companies realize that in order to earn more revenue, the products or interfaces should meet the people's requirements, using habits and other characteristics in the target market. In order to capture the target market, the products or interfaces should be designed and tested for the target people. Rose [5] found that the earlier the localization is considered, the better effects and lower costs will be achieved over the usability life cycles. In a usability life cycle, one of the important phases is to evaluate and modify the product/ interface through iterative evaluation techniques [6]. Previous

studies find that culture influences not only the interface design, but also the evaluation techniques which are used to build the interfaces [7-9]. Since the thinking aloud usability test is considered by Nielsen as the single most valuable usability engineering methods for evaluating the usability of user interfaces [10, p 195], in my study, I will investigate how this method is used in cross-cultural settings.

Think aloud is used as a usability evaluation method to gain insight into how users work with a product or interface. It requires representative users to verbalize their thoughts while performing a task using the system [10, p195]. The primary goal of a usability test is to find a list of usability problems from evaluators' observations and analysis of users' verbal and non-verbal behavior. However, in the cross-cultural usability tests, since evaluators and users come from different cultures, they may be strongly influenced by their local cultural perspective, perception and cognition. Whether the foreign evaluators can get the target users' real feelings and meanings and whether they can find the real usability problems are worth to investigate.

Even though today the common approach to carrying out usability tests in a foreign country is to recruit local evaluators, in some situations, we need to use foreign evaluators instead of local ones. The reasons are:

1. It may be more efficient to use the foreign evaluators, especially in the situation of testing prototypes with target users in order to get quick feedback to the developers. For example, a research team in a company is trying to develop an interface and wants to know whether it is good for people from different cultures. It may be good to consider the "discount usability engineering" [10, p 17] which can be used to improve the interface in a fast and cheap way by using their own evaluators to do the tests with different users. Before an interface being finally implemented, it needs many evaluation life cycles [6]. Usability evaluation is necessary in every application developing phase. Introducing how to do the tests, clarifying the research purpose and the product, training the local evaluator, and waiting for a final report will take a long time [11, 12]. If the evaluators could communicate with the developers frequently while they are doing the tests, it will be more efficient for the developers to modify the prototype. So it is efficient to use the evaluators from their company.

2. It may be less costly to use a foreign evaluator in some situation. For example, a Chinese company hopes to extend products in Denmark. It may be more cost effective if using their own usability professionals to do the test in Denmark instead of employing a Danish usability professional.

3. It may be more effective to use the foreign evaluators when it is hard to get a local expert evaluator.

Considering the cross-cultural cost-benefit analyses [6], using a foreign evaluator happens in the usability engineering area. However, because of the cultural difference, whether the foreign evaluator could find the relevant usability problems is not so clear. Since thinking aloud usability testing is not simply a form of action performed by the user alone, "but a mode of interaction" between users and evaluators [13,p.267], the foreign evaluators themselves might be a potential influence on the users' behavior. Nørgaard's research found that evaluators asked questions about nonexistent parts of the system, speculative or hypothetical questions [14], which implies that the evaluator's feelings and thoughts of the system may call the user's attention or guide the user's direction. This study will mainly focus on investigating how evaluators find usability problems and whether there is any difference between the local and foreign evaluators' problem finding.

Finding the usability problems in the user interface that "could result in human error, terminate the interaction, and lead to frustration on the part of the user" [15, p 247] is one of the most important goals and purposes of the usability testing. In thinking aloud usability tests, usually usability problems are found by evaluators through observing the user to perform tasks, listening to the user's "think aloud", and communicating with the user to elicit more thoughts and opinions. We could easily see that the problem finding involves the interaction between the evaluator and the user and both of them play an important role in finding usability problems. However, since the evaluator is the one who comes up with the usability problems, we only focus on the evaluator's problem finding behavior. Comparing the usability problems found by local and foreign evaluators could give us information about whether they could "read" the users right and interact effectively with the users in the tests.

This study is primarily based on Nisbett's cultural theory [16-19], since it is very relevant to usability tests. This theory discusses the cognition and perception difference between Western and East Asian people. Thinking aloud usability testing is considered as a cognitive activity [20], so it is more appropriate to use Nisbett's theory as the theoretical basis of culture. Thinking aloud processes involve users' cognition and perception characteristics, and the results of the usability test, i.e., usability problems which are found by the evaluators involve the evaluators' cognition and perception of the whole test process. If the evaluator and user are from different cultures, they may be influenced by their local cultural perception and cognition. In this study, we will investigate thinking aloud usability testing in Denmark and China. Denmark is a typical Western country which can represent the society where the usability testing method is developed, whereas China is a typical East Asian country and can also represent those countries that taking the usability testing technique from the Western countries without much changing. So we would like to investigate how Danish and Chinese evaluators find usability

problems with local and foreign users in culturally localized applications.

2. CULTURAL THEORY AND THINKING ALOUD USABILITY TESTING

2.1 Nisbett's Cultural Theory and Hong's Dynamic Constructivist Approach to Culture

Nisbett's cultural theory [16-18] discusses the cognition and perception differences between the Westerners and East Asians, such as field independence vs. field dependence, attention to focal objects vs. attention to field, causal attribution, categorization, logic versus dialectics, etc. These differences are derived from two cognitive styles, analytic vs. holistic cognition and perception [19, 21]. The two cognitive styles indicate East Asians as holistic, dialectical information processing, and Westerners as an analytical, linear thinking style [17, 19, 22]. The Western way of thinking is characterized as analytical, meaning that the people tend to "think in a line", whereas the East Asian way of thinking seems to be more holistic in that they tend to "think in a circle." Previous research [23] found that thinking aloud is best suited to analytical cognitive tasks, while holistic tasks are more difficult to verbalize. The reason may be that when East Asians, who tend to adopt holistic thinking, want to grasp the gestalt of the part, many elements will be held in thought at the same time. It will make the verbalization more difficult to do. In contrast, when Western people, who tend to adopt analytical thinking, break up the object into component elements, it makes the verbalization easier to do [23]. Thus, in thinking aloud usability testing, it may be easier to follow Western users' thoughts than East Asian users'. In order to get the East Asians' thoughts, evaluators need to interact with the users more sufficiently, comparing to the interaction with Westerners.

According to Boren and Ramey [13] who proposed speech communication theory as the theoretical basis for thinking aloud usability tests, few usability professionals, no matter East Asians or Westerners, follow the classical thinking aloud model described by Ericsson and Simon [24] which emphasizes the person who is doing thinking aloud should not be disturbed. Thinking aloud data which is brought out spontaneously by the users is usually incomplete, unclear or ambiguous, so the evaluators need to use undirected and undisturbed tokens or probing technique to make the user stay focused on the tasks and verbalize their thoughts fluently [13, 25]. Although all the users may need the responses from the evaluators, it may be more important for East Asian users. If the evaluators did not realize the difference between the Westerners and East Asians and conducted the tests in the same way with different users, they may have some problems in finding the relevant usability problems for the target user group.

Furthermore, in a cross-cultural usability test, users may behave differently towards foreign and local evaluators. In Nisbett's cultural theory, there are two types of relational orientation: task-focus orientation and socio-emotional orientation [26]. Task-focus orientation means that people's effort is directed towards task-related goals, and attention is focused on monitoring the extent to which these goals are being accomplished. Socio-emotional orientation means that people's effort and attention are focused on the interpersonal climate of the situation, and they strive to maintain social harmony. So users from different cultures may not be influenced to the same degree when they are with a foreign

evaluator. The northern European culture is typically a task-focus culture, which implies that users in those countries may not be influenced as much when the evaluator is from another country since they pay more attention to the task, not the evaluator. While the East Asian culture is a socio-emotional relational orientation culture, users in these countries may be influenced more when they are with a foreign evaluator.

Besides Nisbett's cultural theory, Hong's dynamic constructivist approach to culture [27] could also give us some expectations and explanations for the cross-cultural thinking aloud usability testing. According to this approach, a culture's influence is not static, but related to the situation. The situation could influence the effect of culture on cognition, affect and behavior. One of the important situational factors is applicability. People's behavior will be influenced by situational applicability, which means the appropriateness of a given cultural theory/knowledge depends on who the individual is together with. Sharing knowledge of usability problems and coordinating descriptions of usability problems depend on the mutual perception of group belongingness. When users are with a foreign evaluator, they may have some concerns regarding the appropriateness of raising some issues, such as whether it is appropriate to talk a specific cultural issue with the foreign evaluator, which may make their verbalization data different from that with local evaluators. This could affect the foreign evaluator's usability problem finding.

2.2 Previous Research on Culture and Usability Evaluation Techniques

Yeo [9] examined cultural factors that may affect the results of usability evaluation techniques. From his study, power distance is discovered as an important cultural factor that may influence usability testing. He found that a test user who was of higher rank than the experimenter gave more negative comments about the product than the one who was of lower rank than the experimenter. In another study conducted by Yeo [28], he employed three usability assessment techniques: thinking-aloud technique (objective measure), system usability scale (SUS, subjective measure) and interviews (subjective measure). The results of the Usability evaluations were found to be inconsistent. He found that for the less experienced computer users, or for the users who were not familiar with the evaluators, the objective measure and subjective measure were not matched. Even though these users performed poorly on the task, they still showed a positive attitude towards the software in the interview or SUS. Using Nisbett's task-focus orientation and socio-emotional orientation theory to explain: Malaysian culture is a socio-emotional relational orientation culture, so in the usability test the users hope to establish a harmonic relationship with the evaluator, which makes them reluctant to give too many negative comments, especially when they are not familiar with the evaluators or with higher ranked evaluators or when they do not have confidence about their ability of using the software. So in the usability testing, in order to get more relevant usability problems, the evaluators should introduce themselves and the purpose of the tests sufficiently before doing the tests and it may be better to use the evaluators of the same rank or of lower rank than the test users.

The study conducted by Vatrapu and Pérez-Quiñones [8] showed that even in structured interviews, when with foreign interviewers, Indian users were not willing to talk as freely and accurately as when using a local evaluator. Language may not be the key issue, since in this research both interviewers and users could speak

English fluently. His research found that the culture of the interviewer had an effect on the number of usability problems found, on the number of suggestions made, and on the number of positive and negative comments given. Local interviewer (Indian culture) brought more usability problems and made more suggestions than foreign interviewer (Anglo-American).

Lee [7] explored cultural effects on the process and result of the usability evaluation techniques in Netherlands and Korea. They found that in the usability test, Dutch participants criticized the products more actively and they discovered a product's weakness and also its strength more frequently, comparing to Korea participants. Besides, Dutch participants believed that most problems that occurred during the test were due to the problem with the product, whereas Korean participants believed that problems that occurred during the test were due to their mistakes. For the focus group interview, the results showed that Dutch participants actively engaged in a discussion soon after the interview started, whereas Korean participants took a while to start speaking up. In Korean group, participants rarely spoke voluntarily before they were called upon by the moderator. The moderator needed to call on participants constantly and ask more detailed questions to carry on the discussion. On the other hand, Dutch moderator did not have to do much because Dutch participants actively engaged in discussion, and some of them even had the tendency to speak too long which required the moderator to control such behavior.

From previous studies about usability evaluation techniques used in different cultures, we could see that the way of doing the test or interview does have specific features in different cultures. If a usability professional is used of doing the tests in their own culture, he/she may have some problems when asking him/her to do the tests with target users in another culture, which can be reflected from the usability problem finding behavior.

3. USABILITY PROBLEMS

As talked above, one of the most important goals and purposes of the usability testing is to find usability problems. The definitions of usability problems are not exactly the same for different researchers, but the main idea is similar. Here we would like to introduce the definition proposed by Hertzum and Jacobsen [20, p 422]: usability problems are "the parts of a system that cause users trouble, slow them down, or fit badly with their preferred way of working." What is a real usability problem? A real usability problem is the one that "predicts a problem that users will experience in their own environment, which affects their progress toward goals and their satisfaction" [29, p 46].

When doing cross-cultural usability testing, evaluators from different cultures may tend to have different concept of usability. The different comprehension of usability may make them focus on different aspects of the application, which will influence the usability problem finding. For example, A study about usability construct [30] showed that Chinese participants used constructs related to security, task types, training, and system issues, whereas Danish participants made more use of constructs traditionally associated with usability (e.g., easy-to-use, intuitive, and liked). Since Chinese and Danish people paid attention on different features of the application, they may find different kinds of usability problems. However, on the other hand, since the evaluators are usually trained with professional knowledge about how to do the tests and extract usability problem, when they are

with the users from the same culture, they may not find quite different usability problems. Even though both the local and foreign evaluators found the similar problem, the way to describe the problem or the severity rating given to the problem may also be different. In this study, we examine not only the problems found by local and foreign evaluators, but also the problem severity rating given by the local and foreign evaluators.

3.1 Shared vs. Unique Usability Problems

Usability problems could be divided according to the problems found by single or more groups of evaluators. Usually it has two categories: shared vs. unique usability problems. Shared usability problems are those detected by more than one group of usability testing, and unique usability problems are those identified by only one group of usability testing [31]. The division of shared and unique problems is often used in analyzing problem finding between different groups of evaluators.

Since shared vs. unique usability problem is divided by the differences in evaluators' problem detection [20], it can also be used to examine the problem finding between evaluators in one group to see the evaluators' agreement of finding usability problems in this group. If a problem is found by most evaluators, it may be a common problem with higher priority, which needs to be fixed by the designers. In a group of homogeneous evaluators, if a problem is found by only one evaluator in a single test, it might be a critical problem [32], but there is also a big chance that it is not important. On the other hand, if there are two groups of two different types of evaluators, such as local and foreign evaluators, a unique usability problem is found by one group of evaluators, not the other, then the reason behind it will be worth to discuss. In this study, shared and unique usability problems will be analyzed between foreign and local evaluators

3.2 Usability Problem Severity

The main purpose for rating the problems' severity is to prioritize them and solve the severe problems first, since it is impossible to solve all the problems because of the limitation of time and money [33]. Hertzum [34] found that half of the severity sum (problem impact) was concentrated in approximately 20% of the problems, which means if solve 20% of the most severe usability problems, the application's assessment will be improved 50%. Severity ratings are usually gathered by asking the usability specialists to rate the severity of each problem. One common approach to rating usability severity is to use a single scale [10, p103]. In this study, the rating scale for the severity of the usability problems is [10, 35]:

1= minor usability problem- fixing this when an opportunity arises

2= important usability problem- important to fix, so should be given high priority

3= critical usability problem- imperative to fix this before the system is put into use

4. METHOD

We carried out experiments to investigate the usability problem found and the severity rating given by local and foreign evaluators in Denmark and China.

4.1 Research Design

Two experiments were carried out in Denmark and China. Since they were the same, we only introduce the experiment in Denmark in detail.

The first experiment was carried out in Denmark with Danish users. The independent variable was the evaluators' cultural backgrounds: foreign evaluators and local evaluators. Based on Nisbett's cultural theory, the foreign evaluators should be from East Asian cultures. In this study, all the foreign evaluators were Chinese.

In order to avoid sampling bias, we needed more than one evaluator in each condition. Accordingly, four foreign evaluators and four local evaluators attended this study and each evaluator did four tests, (16 tests with foreign evaluators and 16 tests with local evaluators in total). All evaluators in this study were experienced usability practitioners who had previously done usability tests. All the local evaluators were Danish people who were born and grew up in Denmark, and all the foreign evaluators were Chinese people who were born and grew up in China and were not familiar with Danish culture and unable to speak Danish language.

The users were local Danish people who could speak English fluently. Since all the foreign evaluators can't speak the language in the target country, in order to make sure the difference is not because of the language, but other deeper factors, all the participants should speak English, no matter with local or foreign evaluators.

It is not hard to select Danish people who can speak English well, because there are many courses in English in the university and many people also speak English often in their work. But it may be an issue for Chinese people. All the Chinese evaluators in this study were good at English and most of them had done usability testing in English before. For Chinese users, we interviewed them in English to see their English skill before they came to the test. All the participants should pass the TOEFL or IELTS, or College English Test Band 6 and good at spoken English.

There may be selection bias if using users who can speak English well. But it is not a big issue comparing to the influence of asking local pairs and distant pairs (foreign evaluator-local user pairs) to speak different languages [36]. In order to make sure the difference we find in this study is from the evaluators' perception and cognition difference, not language difference, it is better to ask all the participants to speak English.

4.2 Application and Tasks

We used culturally localized applications, since localized software could act as a primer to elicit the user's culturally related communication with the evaluator, which may influence the evaluator's problem finding.

We designed a "wedding invitation" application prototype by adding a collection of wedding images and icons to My Collections in Microsoft Word's clipart organizer [36, 37]. Since thinking aloud is often used as a formative evaluation [10] which is done in order to help improve the interface, it is appropriate to use an unfinished prototype as the testing application. The wedding invitation application includes three parts, backgrounds folder, invitation text examples and clipart organizer with images. From the application, the user could choose a background, write

the text and select images and icons to make an invitation letter. The application is in the local language. Since in this study, we would like to investigate the localized application, local language could be a part of the localized features. In the industrial area, if the application is designed for the target users, it is usually in the local language. The language of the interface is definitely important for the local application. In this situation, it can also be tested by foreign evaluators if they are familiar with it. For example, the foreign evaluators' company developed it and the evaluators knew the icons in the interface. They can still do the tests with local users even they do not understand the language of the interface. In this study, the wedding invitation application is actually based on Microsoft Word which is familiar by the evaluators and users. We would like to examine whether the evaluators who are doing tests in English with local users could find usability problems in the localized application.

The users' task was to make a wedding invitation for their own wedding. The task was divided into sub-tasks, such as choosing a background, editing the background, writing text, choosing images, etc. The tasks were given to the user one by one.

4.3 Procedure

The tests were conducted in the target countries, Denmark and China. Most foreign evaluators were invited from abroad for this study, and only one Chinese evaluator who did the tests in Denmark is working in Denmark now. But he was born and grew up in China. He was not able to speak Danish and had never seen a Danish wedding invitation before. It is not a big issue to use a Chinese evaluator who is working in Denmark. Because this study is based on Nisbett's cultural theory, which focusing on the cognition differences, even though the Chinese evaluator is working in Denmark now, he grew up in China and got his bachelor degree in China, which could be assumed that he has already had a mature East Asian cognitive style.

We did the tests in the similar usability testing labs with the software called Morae in Denmark and China. In each country, there were two rooms, one testing room and one observation room. Morae was used to record the whole test sessions. With Morae observer, the researcher could observe the tests in the observation room.

Before the tests, the evaluators were informed that they could communicate with the users during the test as they would normally do. We gave such instruction since the approach to thinking aloud was not consistent with every usability practitioner. Whether the evaluator can communicate with the user during the thinking aloud test is controversial for researchers [24]. But in the industrial area, most usability practitioners communicate with the user when doing usability test [13]. In order to make sure the evaluators did the test in their normal way, it was better to tell them that they could communicate if it was necessary to understand the user's speech and find usability problems.

Besides, the purpose of the tests was also introduced to the evaluators in detail. The researcher told every evaluator that the main task for them was to find the usability problems. What the usability problems included in this test was also explained to the evaluators. The problems included not only the functions of the application and how people used it, but also the content of the cliparts we provided and whether people were satisfied with the application or not. It is very important to let the evaluators know

what the problems are, or else, there will be a big individual difference from different understanding of usability problems.

After each usability test, the evaluators were required to write down the problems they found in the test and also rate the severity of each problem. The users were interviewed by the researchers about speaking English to do the test, their feelings of the application and their feelings of doing the test with foreign evaluators.

4.4 Data Analysis

In this study, we calculated:

1) Usability problem discovery, such as how many problems the evaluators found and what kind of problems they found. The problem discovery was analyzed based on the shared and unique usability problems found by local evaluators and foreign evaluators, in order to examine whether there was any tendency of finding specific types of usability problems by local or distant pairs. If there were to be a tendency by the local or foreign evaluators, then it would be safer to draw conclusions. In this study, we analyzed the usability problems based on the two groups of evaluators, not individual evaluators. Because this study focuses on investigating the difference between local and foreign evaluators' usability problem finding behavior, we can't rely on problem finding or severity rating from any single evaluator [33].

2) Severity of the usability problems found by local and foreign evaluators: minor, important and critical.

5. RESULTS AND DISCUSSION

Now we have finished the tests done by Danish pairs, foreign evaluator-Danish user pairs and Chinese pairs. We have not finished the foreign evaluator-Chinese user pairs. We only did four tests with one Danish evaluator and four Chinese users. So we mainly focus on presenting the results of the tests with Danish users in Denmark, and only use some of the Chinese data to compare and explain the Danish data.

First we need to briefly talk the result of the interview about using English to do the tests. Nine Danish users said it was the same to speak English and their local language to do the tests. 11 Danish users and all the 20 Chinese users said it was similar to do the tests in English and their local language. The major problem for speaking English to do the tests was that the application was in local language, when doing thinking aloud in English, the users had to translate the function in the menu into English, which was not easy for those seldom using English version of Microsoft Word. Another problem was that they could not use too many fancy words to express their feelings in English. However, they all thought they had expressed clearly about their opinions and feelings of the application. In the following sessions, we discuss the findings from this study.

5.1 Usability Problems Found by Local and Foreign Evaluators

When with Danish users, local and foreign evaluators found 404 instances of usability problems totally. We constructed a list of 36 non-overlapping usability problems. Among them, Danish evaluators found 196 instances of usability problems including 35 non-overlapped usability problems and Chinese (foreign) evaluators found 208 instances including 36 non-overlapped usability problems. From the number, we could see that there is no big difference between the problems found by local and

foreign evaluators. The reason may be we did too many tests with different evaluators in each condition. Even though the "magic five" [38, 39] which means five users are necessary to capture 80% of the known usability problems of a system is questioned by researchers [31, 40], the doubts are usually related to the diverse contexts of the testing, such as different users, different test conditions, etc. But in this study, the evaluators used the same protocol for the testing and the users were similar, which means the variation in the test setting was small. The only difference was the evaluators' cultural background. However, in each condition, there were four evaluators, and each evaluator did four tests. The four local (foreign) evaluators may be different, so they could find usability problems with much variation. Here we only calculated the total number and categories of the usability problems found by the four evaluators in each condition. There is a big chance to find a sufficient of usability problems for the two groups of evaluators.

Even though the number of usability problems was similar, the problem finding behavior and the description of some specific problems were different. Some specific problems were only found by foreign evaluators after doing the four tests, whereas if checking the problem found in the first two tests, some common problems could be ignored by the foreign evaluators.

5.1.1 The Discussion of the Unique Problem Found Only by Foreign Evaluators

There were 35 shared usability problems which were found by both local and foreign evaluators when with Danish users. Only one problem was the unique problem which was only found by foreign (Chinese) evaluators and none of the Danish evaluators found this problem after 16 tests. The problem was called "problem of no categorization of images: the images should be categorized into some sub-categories." From Nisbett's cultural theory [17, 19], Chinese people are holistic thinking, which implies that they prefer to get an overview of the images before seeing each specific image. On the other hand, Danish people are analytic thinking and there are not so many images in the application, so it is ok for them to see the images one by one. This finding shows the evaluators' influence on the testing. Even though the usability testing is to get information about the user's feelings, opinions and mental models of the product [38], which seems more objective than heuristic evaluation, actually the evaluator's own opinions and feelings of the application also contribute to the result of the usability problems. However, it does not mean that the evaluators just came up with usability problems by misunderstanding the user's behavior. As Boren and Ramey stated [13, p 267], in thinking aloud usability tests, "talk is not simply a form of action" performed by the user alone, "but a mode of interaction" between users and evaluators. The evaluators could probe questions or use undirected and undisturbed tokens to make the user stay focused on a specific task or specific part of the application [13, 41]. When evaluators and users are from the same culture, they may have similar feelings or opinions on the application, which may result in finding usability problems with less variation after doing a sufficient number of tests. When they are from different cultures, the foreign evaluators may miss some problems if did only one or two tests. But if doing more tests with different foreign evaluators, since they are usability professionals, they are able to detect the user's problems but could also find some problems that may be ignored by the local evaluators. The reason may be when with foreign evaluators, the interaction between the user and evaluator has larger variation which results in finding more usability problems. One of the important reasons

that previous researchers suspect the "magic five" is individual difference between users [31, 40]. Their research is mainly based on heuristic evaluation, but actually this consideration can also be used to explain cross-cultural thinking aloud usability testing. From Boren and Ramey's theory [13, 25], the interaction plays an important role in thinking aloud usability testing, which implies that the "unit" of the evaluator and user's interaction is the one deciding usability problems. If the "unit" variation is smaller, "magic five" theory may work, because even doing more tests with similar "unit", the problems can't be increased more. If the "unit" variation is larger, there will be larger chance to find more different usability problems.

This result was different from previous study [8] which showed local evaluators found more problems than foreign ones when doing interview with Indian users. The reason may be Danish users are task-focus orientation [26], who may be not influenced by doing tests with foreign evaluators. From the interview after the tests, we got that most Danish users focused on the task and behaved the same to local and foreign evaluators. Most Danish users said: "if the foreign evaluator did not understand, he/she would ask, or else, I will expect he/she knew what I said and talk the same thing to both local and foreign evaluators." If the users are similar, then the difference may be from the interaction with the evaluators. From the data of Chinese-Chinese pairs, three out of four Chinese evaluators found the problem of "no categorization of images" and each of them found more than half users having this problem. In contrast, two out of four Chinese evaluators found this problem when with Danish users and each of them only found in one test. It shows that the user is still the one who plays the main role in finding problems, whereas the interaction with the user also contributes to the problem finding. This statement can also be confirmed by the way of finding the problems and by the description of the usability problems.

5.1.2 The Discussion of the Way of Finding the Usability Problems

In the usability tests with Danish users, there is a very common problem found by all the Danish evaluators in most tests and all the Chinese evaluators also found this problem after the first two tests. But the way to find it is different. The problem was called "images in the clipart folder are not personal." Personal image seems important to the Danish participants. All the four Danish evaluators found this problem from the first test, whereas Chinese evaluators tended to consider it as a problem from the second test. From the tests with Chinese users in China, we found that only one Chinese evaluator mentioned one instance of problem similar like this, which was "a tool for designing their own background is needed." This information showed that most Chinese people did not consider the impersonal image was a problem for the application. So for Chinese evaluators, even though the problem was so common for Danish users, it was hard for them to find it in the first test. Of course, since they were usability professional and this was a very common problem for the Danish users, the evaluators were able to notice this problem from the second or third test.

From the above discussion, we could draw a conclusion that for the common problems for the users, both the foreign and local evaluators could find the problem after some tests. For the problem which is uncommon for the users but common for the evaluators, the evaluators' interaction with the users may help them to detect this problem, even though it may be easily ignored

by the users. In this situation, usually evaluators with different perspective of the application play an important role in finding the special problems. Besides the problem of "no categorization of images" discussed above, there was another example in the test settings in China. None of the Chinese evaluators mentioned the problem of "impersonal images and can't add own photos" when with Chinese users. The only foreign (Danish) evaluator who did the four tests with Chinese users found three instances of this problem from two tests. Can we say it is because of the user difference? We don't think so, because none of the 16 tests done by Chinese evaluators with Chinese users found this problem, and only one Danish evaluator did four tests found this problem from two tests. It also happened in the previous example of the problem of "no categorization" in Danish settings.

5.1.3 The Discussion of the Way of Describing the Usability Problems

Even though the foreign and local evaluators may find the same usability problem, the way to describe it could also be different. Here we only extract 36 non-overlapped usability problems. But actually it could be more than that, since the description of each instance of usability problem is not consistent. The process of coming up with the non-overlapped usability problems is difficult [35].

The researcher matched the similar problems according to the user's meaning of the problem and the potential way of solving the problem. For example, all the Danish evaluators found the problem called "images in the clipart folder are not personal" with all the Danish users, and all of them described it as "photographs should not be of strangers", or "the images are not personal", or "would not use clipart photos, but their own photos", which emphasized the content of the pictures. On the other hand, the Chinese evaluators also found this problem when they were doing tests with Danish users, but most of them found this problem from the second test, and usually the first time to write down this problem as "would like to have an easy way to add personal pictures", or "need help to guide users making a background by using their personal pictures", which emphasized the functions of the application. But for the application, no matter emphasizing the function or the content, the users pointed to the same thing which was "the images are not personal and need to have the function of adding personal pictures." It includes both the problem and the potential way to solve it. If the problems were closely related, they would be classified as the same usability problem. The reason why we did not distinguish different problem instances inside a non-overlapped problem is that sometimes they are too hard to distinguish, so it is better to use a more general concept. However, we still could see the problem description differences between the local and foreign evaluators within some of the non-overlapped problems. Of course, the Chinese evaluators would describe it the similar way as the foreign evaluators in the following tests when they found most Danish users said so. When the user first mentioned the images were not personal, the Chinese evaluators tended to ask why because they did not think it was a problem to provide impersonal images before the tests, which could be seen from the Chinese-Chinese pairs' tests described above. Hence, they preferred to describe the problem relating to the function of the application when they detected the problem at the beginning, such as not easy to add personal image.

Another example, a problem is called "the image and the background of the invitation does not match." Most Chinese evaluators, no matter with Danish or Chinese users, described it as "the colors of the image and the background are not matched", or "the images and the background should be matched", which emphasized the "matching." This could be regarded as focusing on the content (color) of the image and the background. But on the other hand, most Danish evaluators described this problem as "can't make the image background transparent", which emphasized the functions of the application.

From the above discussion, we could see that when the evaluators describe the problem focusing on its content and when they describe it considering the "function", depend on their understanding of the problem. If they quite understood the user's feeling, for example, if they also thought it was a problem to put other people's photo on their own wedding invitation or if they also thought it was not good to have the image and background unmatched, they would describe the problem as it is. But if they did not have such thoughts before the test and only got the issue from the user's speech in the tests, they would probe questions to help understand the problem in a deeper level, such as, considering the way of "solving" the problem. Having a function of adding own photos and making the image background transparent are the ways to solve the problems of "need own photos, not others' photo" and "the color of the image background and invitation background are not matched" respectively.

5.2 Severity Rating Analysis

The severity rating analysis was based on the non-overlapped problems in each test. If two instances of usability problem pointed to the same non-overlapped problem, the higher rank would be assigned to the problem in this test. Since there were three severity ranks, Chi-Square test, the nonparametric test in SPSS to compare the observed and expected frequencies, was used to examine whether there was any tendency for giving a specific rank by the evaluators. Table 1 shows the percentage of the usability problem severity ranks given by the evaluators with Danish users.

Table 1. The percentage of problem severity ranks given by the evaluators with Danish users

	Danish evaluators	Chinese evaluators
Percentage of minor UPs	35.19%	29.89%
Percentage of important UPs	33.95%	41.38%
Percentage of critical UPs	30.86%	28.73%
X^2	0.481	5.103
Sig.	0.786	0.078

From table 1, we can see that there was no big difference to give minor, important or critical ranks for Danish evaluators, whereas Chinese evaluators tended to rate the problem as important more than minor or critical. When with Danish users, even though the Chinese evaluators had the tendency of giving the "middle" rank to the usability problems, it was only marginally significant $(0.05<p<0.1)$. But we can expect that if there were more tests, the tendency may be significant.

In order to examine the tendency of rating usability problem severity, the tests with Chinese users could also be considered,

though we have not finished the Danish evaluator-Chinese user tests. Table 2 shows the percentage of severity rating given by the evaluators when with Chinese users.

Table 2. The percentage of problem severity ranks given by the evaluators with Chinese users

	Chinese evaluators	Danish evaluator
Percentage of minor UPs	19.10%	43.75%
Percentage of important UPs	61.80%	33.33%
Percentage of critical UPs	19.10%	22.92%
X^2	64.899	3.125
Sig.	0.000	0.210

When with Chinese users, Chinese evaluators had a clear tendency to rate the problem as "important" more than "minor" or "critical." Only one Danish evaluator did four tests with Chinese users, but the total number of non-overlapped usability problems in each test was 48 which could be regarded as a big sample (above 30), so the result can still give us some information.

Considering the finding of table 1 and table 2, no matter with Danish or Chinese users, Danish evaluators did not show a clear tendency of rating usability problems. From the data, we could see that the more severe of the problems, the smaller percentage it would be. But this tendency is not significant. Since this analysis was based on the non-overlapped problems in each test, not each problem instance, and the higher rank was assigned to the problem if some instances pointed to the same problem, this analysis implied that there were more minor and important problems given by the evaluator than the percentage showed in the table and the percentage of critical problems might be less. In the future study, we could analyze the severity rating of each problem instances given by the evaluators. But actually, it makes more sense to analyze the problem severity based on the non-overlapped problems instead of problem instances, since the problems given to the developers should be non-overlapped problems, not all the problem instances. When deciding the severity of the non-overlapped problems, we used the higher rank of the instance severity as the non-overlapped problems. Take the "impersonal images" problem as an example. One evaluator could write more than one instances that belonged to this problem, such as "photographs should not be strangers" and "would like to use own photos." If given different ranks, such as minor and important respectively, for the problem of "impersonal images", the rank should be important not minor, because: 1) there were more than one problem instances pointing to this problems; 2) one of the problem was ranked as important which could imply this problem was an important one.

For Chinese evaluators, they tended to give the middle rank to the problems, and the tendency was much clearer when with Chinese

users. Using Nisbett's cultural theory to explain [17, 21], East Asians are expected to seek compromise solutions to problems and would like to choose the "middle way" solutions for conflicts because of their dialectic reasoning. This feature is reflected in rating problems' severity. If there were only three ranks, Chinese evaluators preferred to choose the middle one. In order to avoid this tendency, we could consider using even number of ranks, instead of odd number of ranks, such as four severity ranks.

6. CONCLUSION

This paper has investigated how Danish and Chinese evaluators found usability problems and rated the severity of the problems in thinking aloud usability tests in Denmark and China. The study shows that there are differences between local and foreign evaluators' problem finding behavior. Local evaluators could find the common problems from the first test, whereas foreign evaluators needed to do more tests in order to find the problems that were not regarded as a problem in their own culture. In order to find sufficient common usability problems for the target group, foreign evaluators may need to do more tests. From this study, we also find that users play the main role in finding usability problems, and the interaction of the evaluators and users also contributes to finding problems. When the evaluators are from another culture, by considering the application from their cultural perspective and through the interaction with the users, they may be able to find some special problems that the local evaluators could not find. Even though some problems were found by both local and foreign evaluators, the description of the problem could also be different, depending on their understanding of the problem. If it was an obvious problem for them, they preferred to describe it in the content level, or else, they would describe it in the function level. For the problem severity rating behavior, Chinese evaluators preferred to choose the middle rank to rate the problems. In this study, they rated the problems as "important" more than the others. Danish evaluators did not show a clear preference of rating the problems.

In the forthcoming study, we will analyze the Chinese data and compare the problem finding behavior in Denmark and China. We will also investigate how the local and foreign evaluators communicate and interact with the users in order to find the usability problems. Moreover, we could compare the problems found and the severity ranks given by evaluators and users to examine whether the local evaluators could understand the users better than the foreign ones.

8. REFERENCES

[1] Faiola, A. and S.A. Matei, Cultural cognitive style and web design: Beyond a behavioral inquiry into computer-mediated communication. Journal of Computer-Mediated Communication, 2005. 11(1).

[2] Honold, P., Cultural and context: an empirical study for the development of a framework for the elicitation of cultural influence in product usage. International Journal of Human-Computer Interaction, 2000. 12(3&4): p. 327-345.

[3] Rau, P.-L.P., Y.-Y. Choong, and G. Salvendy, A cross cultural study on knowledge representation and structure in human computer interfaces. International Journal of Industrial Ergonomics 2004(34): p. 117-129.

[4] Sun, H., Expanding the Scope of Localization:A Cultural Usability Perspective on Mobile Text Messaging Use in American and Chinese Contexts, in Rensselaer Polytechnic Institute Troy. 2004: New York.

[5] Rose, K. and D. Zuhlke, Localization of user-interface-design for mainland China: Empirical study results and their description in a localization model and language issues, in Evolution of the human-computer interaction, X. Ren and G. Dai, Editors. 2005, Nova Science Publishers: New York. p. 87-101.

[6] Mayhew, D.J. and R.G. Bias, Cost-justifying usability engineering for cross-cultural user interface design, in Usabillity and Internationalization of Information Technology, N. Aykin, Editor. 2005, Lawrence Erlbaum. p. 213-253.

[7] Lee, J.-J. and K.-P. Lee. Cultural differences and design methods for user experience research: Dutch and Korean participants compared. in Proceedings of the 2007 conference on Designing pleasurable products and interfaces 2007. Helsinki, Finland ACM New York, NY, USA

[8] Vatrapu, R. and M.A. Pérez-Quiñones, Culture and Usability Evaluation: The Effects of Culture in Structured Interviews. Journal of Usability Studies, 2006, 2006. 1(4): p. 156-170.

[9] Yeo, A.W., Cultural effects in usability assessment, in CHI 98 conference summary on Human factors in computing systems. 1998, ACM Press: Los Angeles, California, United States.

[10] Nielsen, J., Usability Engineering[M]. 1993: New Jersey: AP Professional.

[11] Dray, S.M. and D.A. Siegel, "Sunday in Shanghai, Monday in Madrid" Key Issues and Desisions in Planning International User Studies, in Usabillity and Internationalization of Information Technology, N. Aykin, Editor. 2005, Lawrence Erlbaum: London. p. 189-212.

[12] Dumas, J.S. and J.C. Redish, A Practical Guide to Usability Testing. Revised Edition ed. 1999, Portland, Oregon, USA: Intellect.

[13] Boren, M.T. and J. Ramey, Thinking aloud: Reconciling theory and practice. IEEE Transactions on Professional Communication, 2000. 43(3): p. 261-278.

[14] Nørgaard, M. and K. Hornbæk. What do usability evaluators do in practice?: an explorative study of think-aloud testing in Proceedings of the6th ACM conference on Designing Interactive systems, 209 - 218 2006. USA.

[15] Norman, K.L. and E. Panizzi, Levels of automation and user participation in usability testing. Interacting with Computers, 2006. 18(2006): p. 246-264.

[16] Nisbett, R.E. Cognition and Perception East and West. in 28th International Congress of Psychology in Beijing. 2004.

[17] Nisbett, R.E., The Geography of Thought. 2003, London: Nicholas Brealey Publishing.

[18] Nisbett, R.E. and T. Masuda, Cultural and point of view. PNAS 2003. 100(19): p. 11163-11170.

[19] Nisbett, R.E. and Y. Miyamoto, The influence of culture: holistic versus analytic perception. TRENDS in Cognitive Sciences 2005. 9(10).

[20] Hertzum, M. and N.E. Jacobsen, The evaluator effect: A chilling fact about usability evaluation methods. International Journal of Human-Computer Interaction, 2001. 13(4): p. 421-443.

[21] Nisbett, R.E., et al., Culture and Systems of Thought: Holistic Versus Analytic Cognition. Psychological Review 2001. 108(2): p. 291-310.

[22] Lehman, D.R., C.-y. Chiu, and M. Schaller, PSYCHOLOGY AND CULTURE. Annu. Rev. Psychol., 2004. 55: p. 689–714.

[23] Kim, H.S., We talk, therefore we think? A cultural analysis of the effect of talking on thinking. journal of Personality and Social Psychology, 2002. 83(4): p. 828-842.

[24] Ericsson, K.A. and H.A. Simon, Protocol Analysis. Verbal reports as data. A Bradford Book. 1993: Cambridge Massachusetts.

[25] Tamler, H., High-tech versus high-touch: The limits of automation in diagnostic usability testing. User Experience, 2003. 2(4): p. 18-22.

[26] Sanchez-Burks, J., R.E. Nisbett, and O. Ybarra, Cultural Styles, Relational Schemas and Prejudice Against Outgroups. 2000, University of Michigan.

[27] Hong, Y.-y. and L.M. Mallorie, A dynamic constructivist approach to culture: Lessons learned from personality psychology. Journal of Research in Personality, 2004. 38: p. 59–67.

[28] Yeo, A.W. Global-software Development Lifecycle: An Exploratory Study. in CHI 2001. 2001.

[29] Wilson, C.E., The Problem With Usability Problems: Context Is Critical. Interactions, 2007. 14 (5): p. 46-47, 50.

[30] Hertzum, M., et al. Usability Constructs : A Cross-Cultural Study of How Users and Developers Experience Their Use of Information Systems. in HCI International. 2007. Beijing, China.

[31] Law, E.L.-C. and E.T. Hvanneberg. Analysis of Combinatorial User Effects in International Usability Tests. in CHI 2004. 2004. Vienna, Austria

[32] MOLICH, R., et al., Comparative usability evaluation. BEHAVIOUR & INFORMATION TECHNOLOGY 2004. 23(1): p. 65–74.

[33] Hassenzahl, M., Prioritizing usability problems: data-driven and judgement-driven severity estimates. BEHAVIOUR & INFORMATION TECHNOLOGY,, 2000. 19(1): p. 29-42.

[34] Hertzum, M., Problem Prioritization in usability evaluation: from severity assessments toward impact on design. International Journal of Human-Computer Interaction, 2006. 21(2): p. 125-146.

[35] Hornbæk, K. and E. Frøkjær. Comparing usability problems and redesign proposals as input to practical systems development. in CHI. 2005. Portland, Oregon, USA: ACM.

[36] Sun, X. and Q. Shi. Language issues in cross cultural usability testing: a pilot study in China. in HCI International 2007. 2007. Beijing.

[37] Clemmensen, T. and S. Goyal, Cross cultural usability testing. Working paper, Copenhagen Business School, Department of Informatics, HCI research group, 2005. 2005-006.

[38] Law, L.-C. and E.T. Hvanneberg. Complementarity and convergence of heuristic evaluation and usability test: a case study of universal brokerage platform. in NordiCHI 2002. 2002. Aarhus, Denmark: .

[39] Nielsen, J. and T.K. Landauer. A mathematical model of the finding of usability problems. in INTERCHI'93. 1993. Amsterdam, The Netherlands.

[40] Woolrych, A. and G. Cockton, Why and When Five Test Users aren't Enough. http://www.netraker.com/nrinfo/research/FiveUsers.pdf.

[41] Shi, Q. and T. Clemmensen. Relationship Model in Cultural Usability Testing. in HCI International, 2007. 2007. Beijing.

Different Cultures' Perception of Personas Descriptions

Lene Nielsen
Snitker & Co
Bredgade 21B
1260 Copenhagen K
+45 28707763

ln@snitker.com

ABSTRACT

Personas are considered a method to communicate data on users and to aid the perception of users. Instead of project participants having individual images, the method creates a shared perception of the users that is not built on preconceived ideas, but on field data. This article investigates how different users from different cultures perceive the same textual description of a persona. In the study the participants were asked to find a photo that could illustrate a persona description and explain the choice. The study shows that two strategies seem in play when choosing a photo that can support the description. Either the reader focuses on actual information in the description or the reader interprets the text and uses this as a basis for choosing the right photo.

The choice seems to be related to the informant's gender and age e.g. the only female participating chose a photo of a woman.

Unintentionally, small cues are given in the text these are interpreted from the informant's cultural background.

There seems to be a global stereotypical image of a businessperson, as all informants picked a photo of a person in suit, regarding gender and ethnical appearance.

Categories and Subject Descriptors

H.5.m: Information interfaces and presentation (e.g., HCI), Miscellaneous.

General Terms

Design, Experimentation.

Keywords

Personas, Scenarios, Culture

1. INTRODUCTION

A persona is fictitious user constructed from different forms of field data. The data can originate from a combination of questionnaires, interviews, observations, probes etc. The designer uses the persona to imagine the end user's design preferences and to imagine the needs that the future design can solve. This is done by writing stories – scenarios – about how the persona uses the future system. This enables the designer to explore future possibilities in an easily accessible way and in a format that is easy to change.

A personas description is not just any kind of document, but differentiates from other documents, as it is a description that is aimed at a known reader (the designer), who is distinctively different from the persona it portrays. The portrait will evoke identification in the reader, so the reader can imagine and understand the persona from the description and can, with this method, make informed design decisions.

The literature on the persona method is scant, there are huge numbers of articles on the internet that describes cases, but thorough descriptions of the methods are only found in few works. These describe the reason for using personas within 4 parameters; user focus, communication, design, and market focus:

1. *User focus*. A method to focus on specific users [9]:
2. *Design:* A way to make assumption explicit [9], make informed design choices [6], measure designs effectiveness [2], [6], determine what a product should do [2], engage the design team [9], make better decisions [6], and investigate design ideas [7]
3. *Communication:* Communicate with stakeholders and designers [2], encourage consensus [6], and communicate data [7]
4. *Market*: Contribute to marketing and strategy. [2].

Common to the literature is how the persona is described. The outline of the persona is a written description accompanied by a photo of the imagined user. The writing can have the form of a description, a bulleted list that highlights certain criteria possessed by the user (age, sex, occupation, life situation etc.), posters portraying the persona in typical situations with fictive sentences describing the persona etc.

This experiment originates from a comment posted on the forum connected to Journal of the HCI Vistas[1]. Here Dinesh Katre wrote; '*I have always found it difficult to visualize or understand the characters illustrated in the books of P. G. Woodhouse because all are British personalities and I have not lived in Briton so long to understand these personalities as they are quite culture specific.*' (29-01-08)

From this a simple search on Google scholar for literature on personas method and perception began. The search showed no articles on the subject. A similar search for personas method and culture gives several hits on how the persona method can affect organizational culture, hits on how to consider data from different cultures, but none that takes into consideration how different cultures perceive personas descriptions and what to take into

---

[1] http://www.hceye.org/UsabilityInsights/

consideration when developing personas perceived by readers from different cultures.

2. DIFFERENCES IN CULTURE

In recent years cultural aspects has come to play a significant role when discussion interface design and evaluation methods. In the following I will briefly present the work of the cognitive psychologist Nisbett [8] who argues that there is a difference between how people perceive objects and situations related to the region from which they originate. Nisbett argues that Easterners (Chinese, Koreans, and Japanese) tend to think holistic, are more likely to attend to backgrounds, are more likely to expect change than Westerners, are more likely to group objects in thematic relations, and deal with contradictions finding truth in both sides. Westerners (Europeans and Americans) think analytic, are more likely to attend to objects, group according to taxonomies, and tend to reject one side of contradictions.

In this experiment it is especially the holistic versus the analytic thinking and grouping in themes rather than according to taxonomies that might be in play.

3. PERCEIVING A TEXT AND A STRANGER

The next step was to look at how we as humans perceive others when we have a first encounter and how we perceive a text. In the following I will briefly describe theoretical approaches to perception.

3.1 Perceiving a text

As mentioned before there is consensus on that a persona is presented as a text with an accompanying illustration of the persona either as a photo or a drawing. Personas presentations have commonalities with stories. It applies to stories, that no story can ever be told in its entirety and the reader will, when trying to make meaning of the story, fill in the gaps in the text, known as narrative gaps [4]. This goes for the persona description as well - it is not able to present all information about the persona and the reader infers the missing information. The information that is not received as direct story elements, the reader infers from expectations, knowledge of the depicted area, and own cultural background [1]. Storytelling is a common experience, but the reception is individual and each reader creates his or her individual story. In connection to personas descriptions the question becomes: do culturally different readers fill the narrative gaps differently?

3.2 Encounters - empathic or stereotyped

Similar to the individual filling of narrative gaps based on own experiences, we all have individual perceptions of other people based on earlier experiences. When we meet a stranger, we do not see the person as possessing a unique constellation of characteristics, but add the person to a previously formed category [5] – a stereotype. This category is built on knowledge of meetings with others. Later on, when an in-depth knowledge of the person is formed, the category is broken and the stereotype transforms into a personal character. Stereotypes can be defines as "socially constructed representations of categories of people" [3]. In connection to personas descriptions the question becomes; how different are the categories we use to perceive a description?

4. THE EXPERIMENT

I took a persona description of an online manager from a business project for the Danish company Traceworks. It was translated into English and all mentions of name and living place were removed (see app. 1). From my personal network I asked students and professionals to do as follows:

1. Read the personas description presented below

2. Give the persona a name

3. On the Internet, find a photo that matches the description

4. Describe why you have chosen the photo

5. Before the 1st of June return the name, the photo and the explanation directly to me.

From each of the countries India, China and Denmark I got three replies: five male and three women, all familiar with the personas method. The participants are distributed evenly between students and professionals. In the following I will analyze the response.

5. ANALYSIS

5.1 Strategies

In the accounts for the choice of photo there seems to be two strategies at play; either interpreting or looking for clues in the description:

- Interpreting the text and using the interpretation as explanation for the choice of photo *'I don't know why but I tend to associate obesity with reluctance to new technology mindset. Don't ask me why!'*(Indian informant 3). *'Information Technology means he would earn more money than average'* (Chinese informant 1).

- Finding a specific description in the text that serves as explanation for the choice of photo. *'She has young children and therefore she could not be old'* (Danish informant 2). *'he has no time to do much exercise, so he may be a little fat'* (Chinese informant 2).

Table 1 below shows each of the participants chosen strategy. The Danish informant 1 chose to describe the choice with a personal knowledge of people employed in advertising: *'I assume I know the person, because of my previous career and involvement in advertising. I, from the get go, had an idea of how I presumed this person to look, since to me we all looked alike at the agency after a while.'* (Danish informant 1).

Tabel 1. Perception strategies

Perception based on interpretation	Perception based on description	Perception based on former knowledge
China 1, China 3, India 3	China 2, DK 2, DK 3, India 2, India 3	Dk 1

Most participants used the strategy of finding cues in the text. Interestingly none of the Danes used interpretation as an overall strategy.

Almost all of the photos showed a person in suit, even the only photo of a female. Three from India and China chose photos of western business people, while the rest chose photos of locals.

The Indian informant 1 chose both a westerner and a local looking and explained that in the text is mentioned: 'X is 42 and married, with a son age 3 and a daughter age 7' an Indian man would not have so young children at 42, this points to a westerner.

The photos from two of the Danes looked very Danish while the photo from the Danish female shows a woman, not very Danish in expression. All of the Danes picked photos of younger people than the other informants did.

Tabel 2. Photo contents

Photos of foreign culture	Photos of local culture
China 1, 2, India (1) 2, DK 2	China 3, India 1, 3, DK 1, 2, 3

Tabel 3. Photos arranged according to informant's country

None of the Danes interpreted the text, but rather found cues in the text. Not that this experiment have any statistical significance, but it might correlate to Nisbetts [8] findings of Easterners more prone to holistic thinking than Westerners.

Other areas seem to create a shared understanding, regardless of culture as there seems to be a united, but also stereotypical understanding of how a business person looks.

6. CONCLUSION

This experiment does not show huge cultural differences among the participants. The study is small and further studies are needed to show if personas descriptions are perceived differently in different cultures. Another study with an expectation of larger differences, e.g. a description of a female consumer, might show other findings.

The experiment does show that there are different strategies when perceiving the written descriptions.

The experiment also shows how easily it is to create implicit knowledge in the description, in this case where the age of the persona and the age of children communicates a Westerner.

It also shows how easy it is to create a stereotypical image of the user. As I have written elsewhere, stereotypes prevent engagement [7] and in personas descriptions it is recommended to write 'rounded' character descriptions, but this leads to another discussion.

7. ACKNOWLEDGMENTS

8. APPENDIX 1

The personas description:

X, manager of online marketing.

X is 42 and married, with a son age 3 and a daughter age 7. X lives in a house a bit outside the capital and commutes to and from work.

At work, X likes to wear formal clothes, but not too formal. X tries to do a bit of fitness, but finds it hard to get it into the busy schedule with work and family.

X is content with the place in career and has not, as in the youth, the possibilities to work long hours as X wants to be with the family "I want to play with my children every day". Sometimes X works in the evening, when the kids are in bed.

Education

X has a degree in sales- and marketing from a business school. X studied before there was anything called online marketing, therefore X has the knowledge on online marketing from books and seminars.

The company

X is employed in a large company with a long tradition of offline marketing. X is the manager of the online department in the company and is responsible for the development of the company's online initiatives.

X is responsible for the company's online ads. Here X has to make sure that the campaigns support the company's offline activities. X is also responsible for getting clients for the department's new

activities. X sees a huge benefit in online marketing, as it is easy to access what you get you for your money.

X makes sure that the company's website is constantly updated and that new initiatives make it interesting for the customers to visit the website regularly.

X experiences from time to time that the organization is sluggish and that the road from idea to action can be long.

The online department

The online department employs four people. It is primarily X who is responsible for the company's online marketing. The other employers are mainly engaged in the website and how recently added elements perform. It is important for the department to show a high level of activity on their new initiatives.

Communication

X is in daily contact with the company's many media- and advertising agencies about campaign activities for the next quarter. X lets the media agencies deliver media plans. "I do not have the time to keep an eye on what sites to post on. They are much better at that." X has the final say on the media plans.

X reports directly to the board and presents last quarter's result as well as new initiatives. This is done by collecting numbers from the agency. They are copied into Excel and later into a PPT.

Technology

Apart from X, everybody in the online department is interested in technology. X finds new technologies expensive and difficult to handle, but has a notion that there might be some benefits.

9. REFERENCES

[1] Bordwell, D. 1997. Narration in the Fiction Film. Routledge. London.

[2] Cooper, A., Reimann, R., et al. 2007. About Face 3: The Essentials of Interaction Design. Wiley.

[3] Hinton, P. R. 2000. Stereotypes, Cognition and Culture. Psychology Press. East Sussex.

[4] Iser, W. (2000). The reading process: a phenomenological approach. In Modern Criticism and Theory: a reader. N. Wood. Harlow, Pearson Education: 188-205.

[5] Macrae, N. C. and Bodehausen, G. V. 2001. Social Cognition: Categorical person perception. British Journal of Psychology 92. 239-255.

[6] Mulder, S. and Yaar, Z. 2006. The User Is Always Right: A Practical Guide to Creating and Using Personas for the Web. New Riders Press.

[7] Nielsen, L. 2004. Engaging Personas and Narrative Scenarios. Samfundslitteratur. Copenhagen.

[8] Nisbett, R. E. 2000. The Geography of Thought. Nicholas Brealey Publishing. London.

[9] Pruitt, J. and Adlin, T. 2006. The Persona Lifecycle: Keeping People in Mind Throughout Product Design. Morgan Kaufmann. San Francisco.

Extracting Users' Data: Towards development of a cultural and semantically sensitive combinatorial methodology

Pradeep Yammiyavar
Department of Design, Indian Institute
of Technology, Guwahati, India

pradeep@iitg.ernet.in

Torkil Clemmensen
Informatics, Copenhagen Business
School, Copenhagen, Denmark

tc.inf@cbs.dk

ABSTRACT

Existing techniques of information elicitation and data extraction from the user are successful in defining the user at the surface level. It is argued here that Designers and Usability researchers need data from within the user that are to be mined from deeper cognitive levels of the user. A combination of existing methods has a better capability of doing this deep data mining as against using each technique separately. Two cases, one of them non invasive using eye movements and the other involving retrospective probing through an interview are used as illustrations of possible techniques to be used in combination.

Categories and Subject Descriptors

H5.2 [**Information interfaces and presentation (e.g., HCI)**]:
User Interfaces--*Evaluation/methodology.*

General Terms

Human Factors.

Keywords

Data extraction, cognitive levels, Usability.

1. INTRODUCTION

In Usability Evaluation methods the User is undoubtedly the focus. There are a number of techniques practiced universally to not only observe and study the user but also to extract date that will be useful in Usability evaluations and design of new products / systems. Some of the widely adopted methods are questioners, interviews, think - aloud sessions, focus group interaction, ethnographic studies, action research, and heuristic observations to name a few (Ericsson & Simon, 1993; Nielsen, 1993). Consumer study and marketing literature too has well developed methods of acquiring data of the users from the users themselves(Hayes, 1992; Urban & Hauser, 1993).

Of late it is observed that designers and usability researchers are frequently engaged in elicitation of not only data *'from'* the user but also *'of'* the user and more importantly from *'within'* the user (Yammiyavar, Clemmensen, & Kumar, 2008).

However it is contended here that depending upon the eliciting techniques used, such data gives information at the collective level of the population . In the techniques most commonly adopted, individual users are interviewed , observed and experimented upon.

While there could be demographic commonality of a set of users, the wide differences due to individual variations as a result of say personal attitudes or cultural influences, may not get reflected when the data is collated and analysed. Such data can therefore said to be at the gross or mass or surface level of application to the population as a whole but may not be valid if one were simultaneously interested in mapping it to the individual user's metal model. Such representation of what the user actually semantically means when he or she responds to a interview question or executes assigned tasks in an experimental set up, can get lost due to cultural differences and language nuances.

Another strong reason for looking at new combinations of established user data extraction techniques is the increasing emphasis on the cultural influences. Several researchers have studied and investigated user data elicitation methods for their efficiency, appropriateness, effectiveness and ability to elicit users preferences, actions and behaviors. This will directly affect the quality of the Mental models that a Usability engineer develops and makes use of for determining the interaction protocols & information architecture of a product. Of recent research focus is the role of Culture both in Usability evaluation methods and mental model development. The need for culturally sensitive testing protocols have been widely felt by the software industry.

User's culture has been found to influence the cognitive process of the user by well known and widely quoted researchers such as (Hofstede, 1980; Marcus, 2002; Nisbett, 2003). The socio-cultural settings of the individual also has been found to influence the users behavior in interaction with others The need for accommodating and accounting for culture in the usability testing setups and interface design process have been iterated strongly time and again.

Therefore it is argued here that Designers and Useability researchers need data from within the user that are to be mined from deeper cognitive levels of the user. We define deep level cognitive data as the one influenced by attitudes + (plus) preferences + (plus) Cultural norms + (plus) localised practices. It is posited that there are different levels of needs and wants and desires in users. Psychologists state that there are many variables that govern the individual's reactions. In a number of illustrative design research projects currently being worked upon at the Indian Institute of Technology Guwahati, India, (some of these are done in collaboration with the CULTUSAB project) - methods of extracting deeper level data concerning Semantics, Culture, Emotions, Ethnography as well as their analysis & utilization in designing usability are in progress. Tow of them are outlined bellow as case studies.

2. CASE STUDY 1

Non invasive techniques of extracting Usage data from Users: EMR and the design of dynamic personalized interfaces (Vaishnav & Yammiyavar, 2007).

In this case the design problem was to identify user's preferences and reading behavior of news items while surfing and reading his/her usual daily news paper on an online computer screen . The product that was being designed was an intelligent recommender system that would capture user behavior patterns-unobtrusively through an Eye Movement recorder (EMR). The data captured over a period of two weeks would make the computer intelligently aware of the sub- headings, news topics, advertisements etc that this particular user prefers. These preferences and behaviors were in turn recommend to the computer by an intelligent recommender so as to enable it to display a rearranged lay out on the screen when ever that particular user logged in. This meant that if the user preferred to read Sports news first, then sports news would be on the first page. Similarly only those advertisements that the user wanted to see or was interested in , would be displayed, there by avoiding noise & clutter. The news paper interface would have a continuously dynamically changing graphical lay out depending upon the user's preferences. All this happens unobtrusively without the user being aware of. Data of the user has been collected in a non invasive manner by the EMR. Post session interviews were held and additional data was obtained. Recordings of EMR and videos were played in order to confirm and know why they went to a particular part of a page. A combination of techniques led to better understanding of the user's intentions and actions.

If one were to set out collecting data in a conventional way through questioners etc from a set of users , one would find it difficult to come up with conclusions from such a data that would help decide a page lay out that a majority of the online news paper users found comfortable. Even within a page users have different reading behaviors. Hence the decision of a fixed lay out to a sub group would not necessarily make that sub group happy about the page layout even though the data for the layout was extracted from that sub group.

In the above project a system was designed that intelligently captured the user's reading habits in a non invasive technique and recommended the system to come up with a different Interface depending upon who the user was. Eye movements were recorded using a non physical contact infra red based Eye moment recorder. Mouse clicks were captured in the background. As users went about using the news website according to their own routine habits. The data captured was processed by the EMR software which generated heat maps, (concentrated and prolonged viewing of an area) , gaze plots and videos. The videos were used for supporting post usage interview content analysis. The results indicated the existence of unique behavior patterns in users while using the website. These behavior patterns were used to develop a profile of the user which in turn was used to make recommendations to the user via a software written for the purpose.
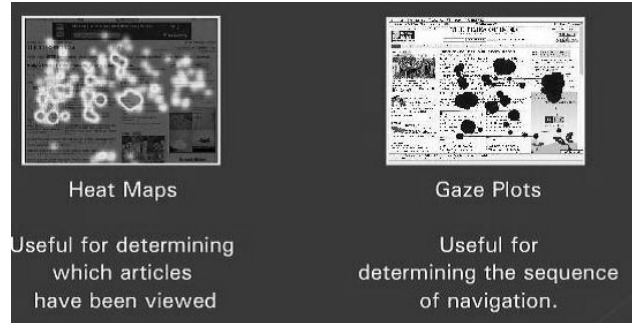

Figure showing three different forms of data collected by experiment.

3. CASE STUDY 2

Retrospective verbalization recording and analysis technique (Kumar, Yammiyavar, & Nielsen, 2007).

Think Aloud, (TA) which is the most prevalent Usability evaluation method in practice, and which is a form of concurrent data collection has been criticized on accounts of interference with task and incompleteness of reported verbal data. One of the prominent criticism of TA is that of sharing the cognitive resources of the user with task fulfillment. TA data collection technique has been reported to interfere with the task especially in cases where the task is cognitively demanding.

An attempt was made to try and capture retrospectively data from the user based on the recollection of the user's experiences with the help of video recordings in a technique named as Mind Tape method. The users are probed and questioned about why's and how's of the on-screen behavior under the stimulus of the replayed screen recording of user activities. The stimulus acts as cue to enable the recall of reasoning behind a particular activity / action / non action during the test. Retrospective replay is adopted with the assumption that, certain secondary and deeper level cognitive processes can be pulled up to the surface without loss or bias of after thought. Under the influence of the stimulus mind acts as a tape and unwinds the memory, thread by thread. By appropriate interviewing, each thread can help trace back deeper level reasoning.

The participants were observed from both Danish as well as Indian cultural backgrounds during the test to be comfortable with the Mind Tape method. No participant reported physiological discomfort due to constant verbalisation which has been observed in TA sessions. Sometimes users also divulged personal information that they were reminded of during the interaction. Indian users participated more voluntarily and divulged more information on their own and provided logical explanations than the Danish participants. Indian users were found to speak more elaborately about what their expectations and outcomes during the interaction were. On the whole, the Mind Tape data was richer in terms of explanations in case of Indian users than Danish user.

Results suggest that not only the sequence of activities but also the intentions and motives of the users behind choices made are traceable using this method. The sensitivity of Mind Tape method in surfacing out the cultural cognitive differences and similarities for direct usage by designers has been observed.

It is posited here that a combination of concurrent technique (example TA) and a retrospective technique (example Mind Tape) is likely to yield much better user data. A combination would help the researcher to link action to deep rooted reasons in the user for those actions.

Other techniques that hold potential as a combinatorial set are the analysis of Gestures (Yammiyavar et al., 2008). Gestures can yield useful data to Usability researchers.

Kelly's Repertory Grid Technique also holds good potential to be used in combination with other techniques. The RGT technique has the advantage that the data elicited is based on the mental constructs of the individual himself. Reference points for these constructs are often entwined with the user's attitudes, beliefs and culture systems. In combination with retrospective techniques it is possible to trace back these reference points that influence the formation of the constructs.

4. CONCLUSION

This paper attempted to put forth suggestions for combining techniques to be able to extract deeper level cognitive data concerning the user. A combinations of existing techniques – of User's data elicitation, that would be much more comprehensive, useful, and sensitive to requirements of semantic determinants such as affect and cultural sensitive has been argued for.

5. ACKNOWLEDGMENTS

6. REFERENCES

1. Ericsson, K. A., & Simon, H. A. (1993). Protocol Analysis. Verbal reports as data: Cambridge Massachusetts.

2. Hayes, B. E. (1992). Measuring Customer Satisfaction: Development and Use of Questionnaires: ASQC Quality Press.

3. Hofstede, G. (1980). Culture's Consequence: Comparing Values, Behaviours, Institutions and Organizations Across Nations

4. Kumar, J., Yammiyavar, P., & Nielsen, J. (2007). MindTape technique - a usability evaluation method for tracing cognitive processes in cross cultural settings e-Minds, 1(3).

5. Marcus, A. (2002). Fast forward: Culture class vs. culture clash. Interactions, 9(3), 25-28.

6. Nielsen, J. (1993). Usability Engineering: Academic Press.

7. Nisbett, R. E. (2003). The Geography of Thought. London: Nicholas Brealey Publishing.

8. Urban, G. L., & Hauser, J. R. (1993). Design and Marketing of New Products, 2. Aufl., Englewood Cliffs.

9. Vaishnav, V., & Yammiyavar, P. G. (2007). Identifying Patterns in User's Preferences and the role of culture behind it – a case study of News Website: IITG Guwahati.

10. Yammiyavar, P. G., Clemmensen, T., & Kumar, J. (2008). Influence of Cultural Background on Non-verbal Communication in a Usability Testing Situation International Journal of Design 2(2), 31-40.

Usability and Culture as Two of the Value Criteria for Evaluating the Artifact
- A New Perspective from the Artifact Development Analysis (ADA) -

Masaaki Kurosu
The Graduate University for Advanced Studies
Hayama-shi, Kanagawa, Japan
National Institute of Multimedia Education
Chiba-shi, Chiba, Japan
masaakikurosu@spa.nifty.com
+81-3-3984-7703

ABSTRACT
In this paper, the conceptual framework of the Artifact Development Analysis (ADA) and its relationship to the usability engineering are outlined. The ADA provides the viewpoint for any artifacts including the hardware, the software, the humanware, the behavioral pattern and the system including them all. Its viewpoint extends both in temporal and spatial dimensions. In short, it deals with the diversity of the artifact and casts the questions "why it is so" and "why it is not so". In this respect, the ADA is related to the usability engineering as one of the value attitudes. The usability engineering puts emphasis on the effectiveness and the efficiency of the artifact. It is not always the value attitude of the highest importance and some people sometimes put more emphasis on other criterion such as the aesthetic aspect, the cost, etc. Based on the findings of ADA, it is possible to see to what extent the usability of some artifact could give the core satisfaction to the user and also to provide the guideline on how the artifact should be designed.

Author Keywords
Artifact development analysis, culture, usability, value system, design, user engineering

USABILITY OF ARTIFACT

What is the Usability?

There are many definitions of usability among which the

one proposed by Nielsen and the other defined in ISO9241-11 are well-known. The former definition of usability is a sub-concept of the usefulness and is a sub-sub-concept of the acceptability. But because Nielsen differentiated the usability from the utility, the usability has a non-negative connotation whereas the utility has a positive connotation. On the other hand, ISO9241-11 proposed a definition of the concept with three sub-concepts namely the effectiveness, the efficiency and the satisfaction. In this definition, both the utility and the usability that Nielsen defined are included in the definition of usability. Hence sometimes the definition by Nielsen is called the small usability and the definition of ISO9241-11 is called the big usability. The ISO definition has been rather famous in Europe and Asia, but recently it has become accepted in North America.

Although the definition of ISO9241-11 is currently accepted world-wide, Kurosu proposed a revised definition of usability that is expressed in Figure 1. Basic ideas are as follows;

(1) The satisfaction is different from the effectiveness and the efficiency because the former is the subjective impression on the side of the user and the latter two are the objective characteristics on the side of the artifact.

(2) The satisfaction is dependent on the effectiveness and the efficiency whereas the latter two are mutually exclusive and are independent with each other.

(3) The satisfaction is dependent on more factors than the effectiveness and the efficiency. Factors include other quality traits such as cost, safety, reliability, compatibility and durability, and some groups of subjective characteristics such as sensibility and emotion, needs and emotion, and value system.

(4) The satisfaction is the ultimate criterion of the artifact.

(5) Hence, the usability engineering should focus on the concept of usability that consists of the effectiveness and the efficiency whereas the user engineering is

focusing on the concept of satisfaction including the usability as a sub-concept.

(6) The artifact should finally be evaluated in terms of the satisfaction. (in other words, the usability evaluation is evaluating just one aspect of the artifact)

efficiently, so that with satisfaction. It is contrary to the natural objects that are intact by the human being. Artifacts include the hardware, the software, the humanware, the behavioral pattern and the system as the integration.
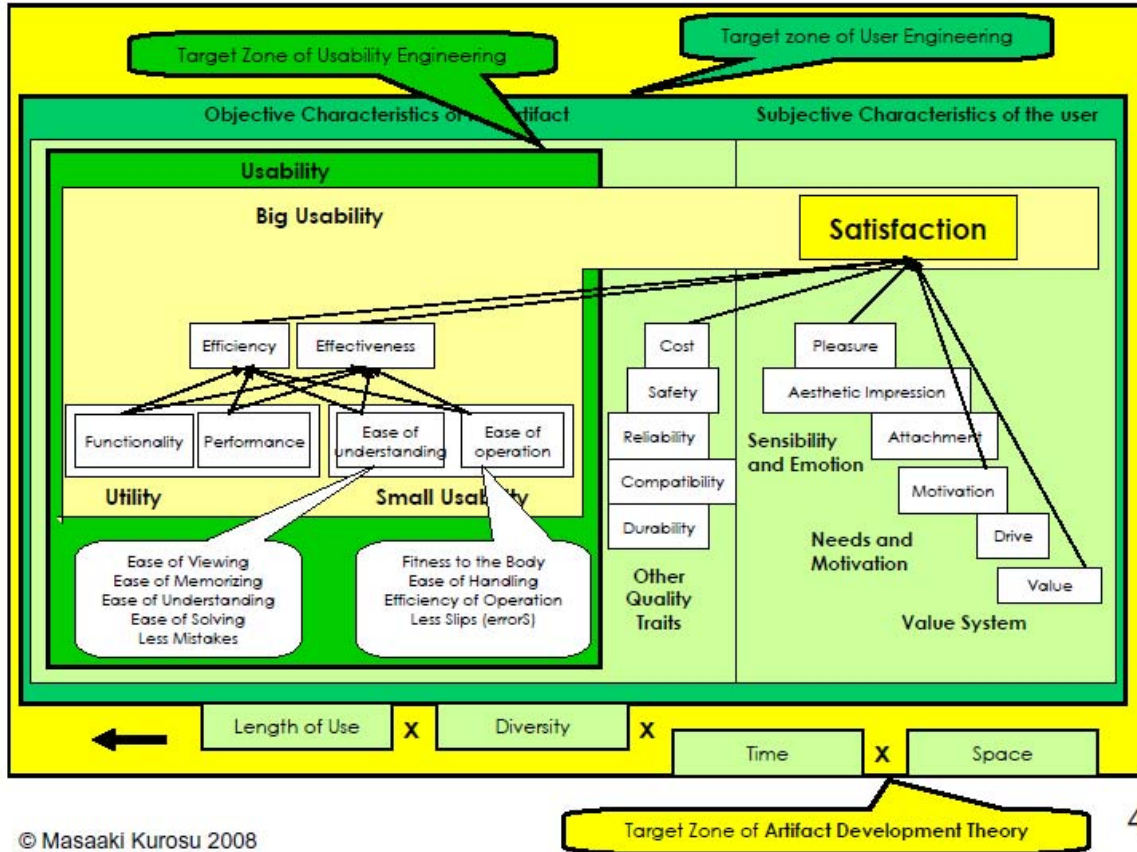
Figure 1. The concept of the usability and the satisfaction

ARTIFACT DEVELOPMENT ANALYSIS (ADA)

What is the Artifact?

From the viewpoint of user engineering, artifacts are invented, designed, and redesigned so that the goal achievement of human being can be facilitated effectively and efficiently. Fundamental scheme of this idea are represented in Figure 2 and 3. In situations where the user can hardly achieve the goal, an artifact is designed so that it facilitates the goal achievement in the right direction and in just the shortest time.

The artifact is an object that the human being created, produced, manufactured, altered, diverted or altered for supporting the goal achievement behavior effectively and
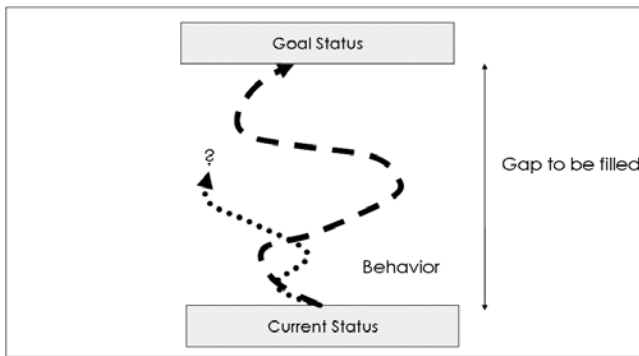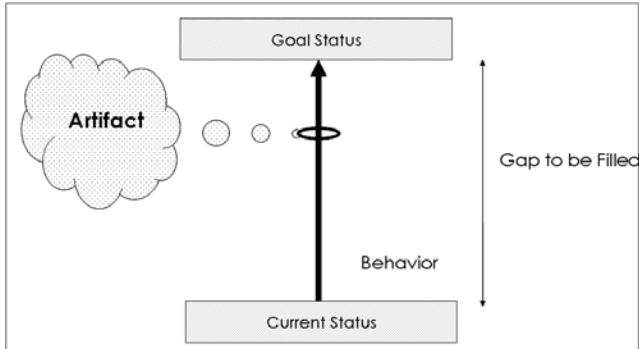
Figure 2. The goal achievement

Figure 3. The goal achievement and the artifact

What is the Artifact Development Analysis (ADA)?

The Artifact Development Analysis is the scientific research approach that analyzes artifacts that were invented and used by people of specific period of time and of specific region for the achievement of some specific goal. It is related to the user engineering but is a science that seeks for the logic underlying the reality whereas the user engineering is an engineering that looks for designing that is satisfactory to the user.

Fundamentally, the ADA takes following stances.

(1) ADA seeks for answers to following questions - What kind of variations are there? Did each of variations have a necessity to be designed as such? Wasn't there a possibility that any different type of artifact should be designed or selected?

(2) ADA evaluates each one of variations to what extent it is reasonably adapted to the goal. Check if there are residual problems.

(3) Finally, ADA specifies whatever artifact is necessary that fulfills the conditions for achieving the goal.

Generally ADA approach takes following steps.

Step 1 - Discover the diversity among artifacts based on the approaches including history, archaeology, cultural anthropology, ethnography and folklore

Step 2 - Find out the commonality and the difference among artifacts

Step 3 - Investigate the reason why and why not

Step 4 - Pursue the inevitability of the specificity of design

Step 5 - Obtain the evaluation for other design to see if the current design is optimal and is acceptable and will give the satisfaction

Step 6 - Consider about the underlying value system that satisfies the user.

Step 7 - Integrate the evaluation from the viewpoint of goal-achievement

Step 8 - Consider if some design that is better than the current artifact is possible or not in the direction of the value system that it is evaluated by the user.

Step 9 - Set up the design guideline

Variation among Artifacts

Variations among artifacts have two dimensions namely the time and the space. The time dimension includes the historical time and the individual time. Thus the ADA approach is related to the history, the archaeology and the psychology. And the special dimension includes the real space and the virtual space where the former includes the geographic space, the political space, the ethnological space and the cultural space and the latter includes the conceptual space and the organizational space. In this sense, the ADA approach is related to the cultural anthropology, the ethnology, the ethnography, the folklore and the sociology.

There are very many factors to generate the diversity of which the ADA focuses attention. Table 1 shows the list of possible factors that may affect the diversity among artifacts.

3

Table 1. Factors Affecting the Diversity among Artifacts

Factors specific to		
the manufacturer	the user	the social group
Availability of the material	Importance of the goal	Persistence to the tradition
Availability of the processing tool	Physical characteristics	Group conformity
Characteristics of the object	Psychological characteristics	Historical background
Manufacturing cost	Social context of use	Ethnic consciousness
Brand power	Physical environment	Influence of the religion
Emphasis on the maintenance	Purchasing ability	Degree of multi-ethnicity
Emphasis on the reliability	Expected life span	
Emphasis on the safety	Literacy	
Aesthetic sense of the designer	Attribution to the social group	
	Sensitivity to the fashion	
	Aesthetic sense of the user	

Variety of Goals

Usually the goal can be expressed as the verb as follows.

Obtain (incl. purchase)
Eat or drink
Preserve
Cook
Record (incl. write)
Inhabit
Sleep
Know the time
Empower the sense (incl. eye-glass, hearing aid)
Communicate
Know the location
Enjoy
Wear
Clean (incl. the body, the clothes, the house, etc.)
Move (incl. walk)
Fight
Punish
Maintain the society

VALUE ATTITUDE AND CULTURE

Value Attitude for Artifact Evaluation

Table 2 is a tentative list of the value attitudes based on the original idea of Spranger. In this table, the usability is just one element of the whole value attitudes. It is the culture that differentiates the weight vector for this value attitude. In other words, there is a culture that emphasizes the usability but there is also another culture that emphasizes the aesthetic impression. Diversity among artifacts that are designed for supporting the same goal achievement can be derived from the difference of value attitudes.

Acceptability of Diversity

Interesting is the fact that some diversity among artifacts can be acceptable while other can not be. The acceptable diversity is the case where each alternative has its own advantage and is used in different environment. The example is the car and the bicycle. There are a few type of unacceptable diversities. One case is that some alternative is evidently better than others. The example is the case of the SD memory card vs. compact flash, smart media, memory stick, xD-picture, MMC, etc. And another case is that some alternative is evidently inferior to others. The example for this case is the current way of enjoying the music that includes downloading of the music from internet, storing the music in the memory card vs. storing the music in the cassette tape, mini disc, or sound sheet. Finally, there is a case where each alternative has its own advantage but the context of use is duplicated. The example for this case is that the numerical key pad for the calculator (IT) and that for the telephone (CT) should now be integrated in the era of ICT though they could exist in different domains in the past.

Model of the User and the Designer

Figure i and j shows the model of actual user and the normative model of designer. Both figures are adopting the value criteria as shown in Table 2.

These figures include the usability and the culture as two of key components and should be good frameworks for the discussion.

CONCLUSION

In this paper, the conceptual framework of the Artifact Development Theory (ADA) and its relationship to the usability engineering were outlined. The ADA provides the analytical viewpoint for the artifact. Its viewpoint extends both in temporal and spatial dimensions. In short, it deals with the diversity of the artifact and casts the questions "why it is so" and "why it is not so".

In this respect, the ADA is related to the usability engineering as one of the value attitudes. The usability engineering puts emphasis on the effectiveness and the efficiency of the artifact. It is important bus is not always the value attitude of the highest importance and some people sometimes put more emphasis on the aesthetic aspect, the cost, etc.

Based on the findings of ADA, it is possible to see to what extent the usability of some artifact could give the core satisfaction to the user and also to provide the guideline on how the artifact should be designed.

Table 2. List of Value Attitudes That Are Related to the Evaluation of the Artifacts

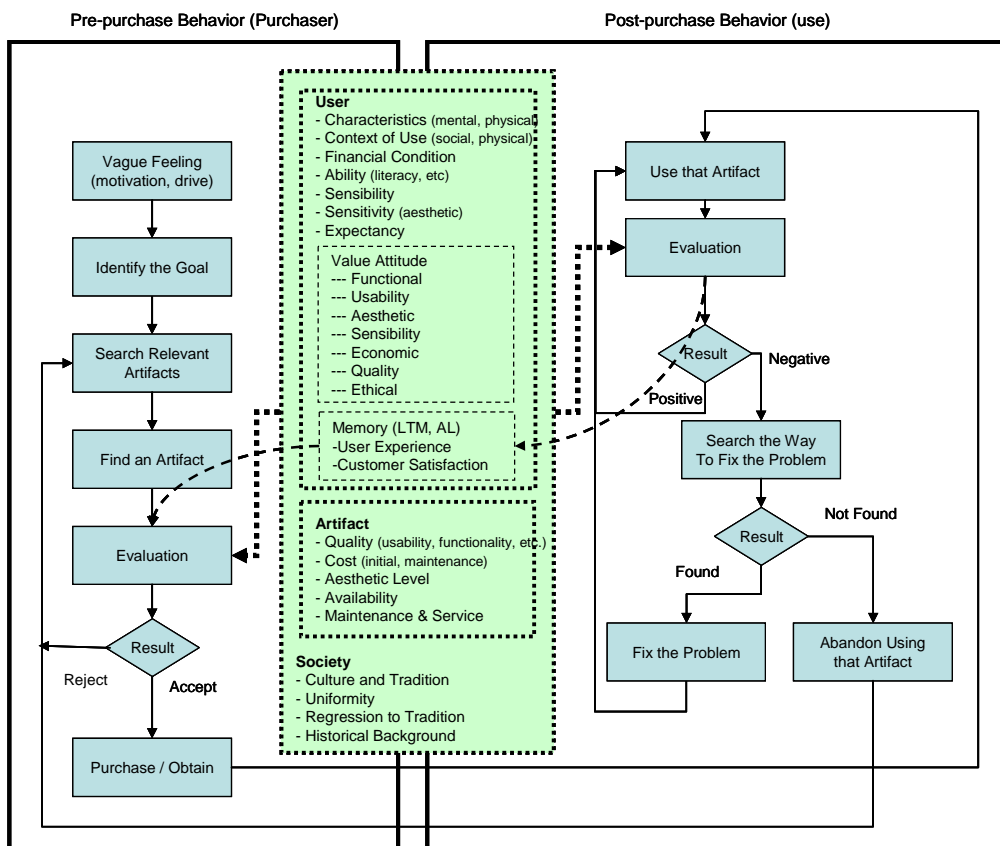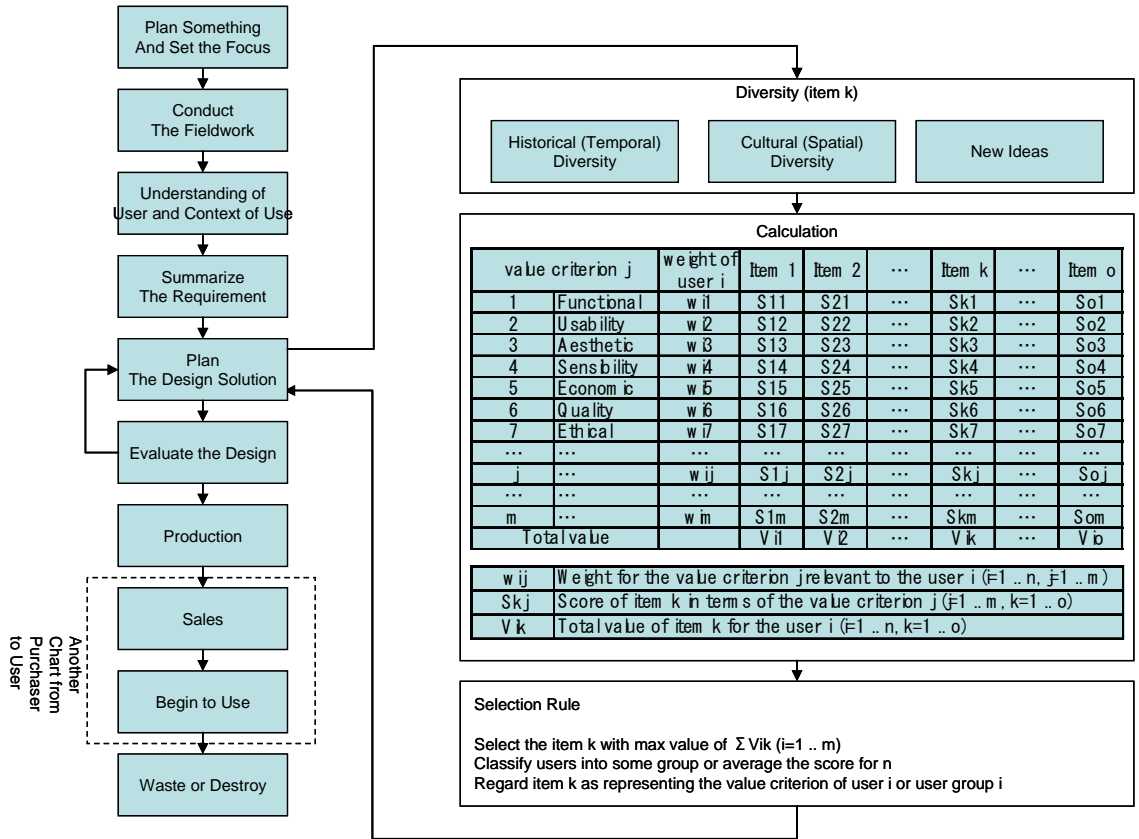|	
Functional Value Attitude	Put emphasis on a new function and/or the multi-functionality
Usability Value Attitude	Put emphasis on the effectiveness and the efficiency
Aesthetic Value Attitude	Put emphasis on the appearance and the good-looking design
Sensibility Value Attitude	Put emphasis on the attachment or the emotional relationship
Economic Value Attitude	Put emphasis on the cost (initial cost and maintenance cost)
Quality Value Attitude	Put emphasis on the qualities such as the reliability, the safety, and the compatibility
Ethical Value Attitude	Put emphasis on the emvironmental aspect and the sustainabillity

Figure i. User's model (actual) based on the ADA

Plan Something And Set the Focus

Conduct The Fieldwork

Understanding of User and Context of Use

Summarize The Requirement

Plan The Design Solution

Evaluate the Design

Production

Sales

Begin to Use

Waste or Destroy

Another Chart from Purchaser to User

Diversity (item k)

Historical (Temporal) Diversity | Cultural (Spatial) Diversity | New Ideas

Calculation

value criterion j		weight of user i	Item 1	Item 2	...	Item k	...	Item o
1	Functional	w_{i1}	S_{11}	S_{21}	...	S_{k1}	...	S_{o1}
2	Usability	w_{i2}	S_{12}	S_{22}	...	S_{k2}	...	S_{o2}
3	Aesthetic	w_{i3}	S_{13}	S_{23}	...	S_{k3}	...	S_{o3}
4	Sensibility	w_{i4}	S_{14}	S_{24}	...	S_{k4}	...	S_{o4}
5	Economic	w_{i5}	S_{15}	S_{25}	...	S_{k5}	...	S_{o5}
6	Quality	w_{i6}	S_{16}	S_{26}	...	S_{k6}	...	S_{o6}
7	Ethical	w_{i7}	S_{17}	S_{27}	...	S_{k7}	...	S_{o7}
...	...	...	...	...	...	...	...	...
j	...	w_{ij}	S_{1j}	S_{2j}	...	S_{kj}	...	S_{oj}
...	...	...	...	...	...	...	...	...
m	...	w_{im}	S_{1m}	S_{2m}	...	S_{km}	...	S_{om}
Total value			V_{i1}	V_{i2}	...	V_{ik}	...	V_{io}

w_{ij}	Weight for the value criterion j relevant to the user i ($i=1 .. n$, $j=1 .. m$)
S_{kj}	Score of item k in terms of the value criterion j ($j=1 .. m$, $k=1 .. o$)
V_{ik}	Total value of item k for the user i ($i=1 .. n$, $k=1 .. o$)

Selection Rule

Select the item k with max value of ΣV_{ik} ($i=1 .. m$)
Classify users into some group or average the score for n
Regard item k as representing the value criterion of user i or user group i

Figure j. Designer's model (normative) based on the ADA

REFERENCES
1. Lewis, E.St.E. (1908) "Financial Advertising", Reprint: N.Y. & London: Garland Pub. (1985)

2. Sheldon, A.F. (1911) "The Art of Selling" Chicago: The Sheldon School

3. Howard, J.A. and Sheth, J.N. (1969) "The Theory of Buyer Behavior" Wiley

4. Engel, J.F., Blackwell, R.D. and Kollat, D.T. (1978) "Consumer Behavior" Holt Rinehart and Winston

5. Blackwell, R.D., Engel, J.F. and Miniard, P.W. (2005) "Consumer Behaviour" South Western, Div of Thomson Learing International

6. Spranger, E. (1914) "Lebensformen" Halle: Niemeyer (translated by Pigors, P.J.W. NY: G..E.Stechert Co.)

7. Asai (1974) "A research on the value attitude questionnaire 2" (in Japanese) Proceedings of Japanese Educational Psychology Organization P.510-511

8. Sakai (2001) "Individual difference for the value concept and the background: A study based on the value scale" (in Japanese) Proceedings of Japanese Educational Psychology 49(1) P.102-111

9. Riesman, D. (1961) "Culture and Social Character: The Work of David Riesman Reviewed" Free Press

10. Riesman, D. (1969) "The Lonely Crowd: A Study of the Changing American Character" Yale Univ. Press

11. Kurosu, M. (2007) "A proposal for the Artifact Development Analysis" (in Japanese) Journal of Human Centered Design Vol.3, No.1

12. Kurosu, M. (2007) "An approach to the Artifact Development Analysis" (in Japanese) Human Interface Symposium 2007

13. Kurosu, M. (2007) "Structure of artifact development for the achievement of goal" (in Japanese) Proceedings of Japanese Psychological Association 2007

14. Kurosu, M. (2007) "Optimality of design from the scope of artifact development" (in Japanese) SIG USE of Human Interface Society

15. Kurosu, M. and Hashizume, A. (2008) "Selection of the communication media and its dependency on the situation" (in Japanese) Journal of Human Centered Design Vol.4, No.1.

16. Kurosu , M. and Ando, M. (2008) "Non-selection and waste of artifact" (in Japanese) Proceedings of Japanese Psychological Association.

7

Cultural Usability in Computer Supported Collaboration

Ravi Vatrapu

Hawai'i Computer-Human Interaction Laboratory (hi'CHI)
Department of Information and Computer Sciences
University of Hawai'i at Mānoa
POST 317, 1680 East-West RD, Honolulu, HI, 96822 USA.
vatrapu@Hawai'i .edu

ABSTRACT

This paper presents an empirical analysis of usability in the emerging field of computer supported intercultural collaboration (CSIC). The basic premise of the research project on which the empirical analysis is based is that social affordances of technologies vary across cultures. To empirically evaluate the premise, an experimental study was conducted to investigate how pairs of participants from similar and different cultures (American-American, American-Chinese, and Chinese-Chinese) appropriate affordances and relate to each other in a computer supported collaborative learning environment. Usability analysis consisted of evaluating objective performance measures and subjective user interface satisfaction measures. Statistical results show a systemic variation between cultures. We discuss the implications of these findings for the research and practice of cultural usability.

Categories and Subject Descriptors

H.5.3 Group and Organization Interfaces: *Theory and models, Asynchronous interaction Collaborative computing, Evaluation/methodology;* H.1.2 User/Machine Systems: *Software Psychology*.

General Terms

Affordances, appropriation, culture, cultural usability, technological intersubjectivity.

1. INTRODUCTION

Cultural aspects of usability have been a topic of study in the field of human-computer interaction (HCI). Early research focused on localization and internationalization user interface aesthetic issues of languages; colors; and conventions of data, time and currency [15, 23, 30]. Subsequent research investigated cultural considerations in the usability evaluation methods employed and usability problems detected. Prior empirical work has documented cultural differences in the *functioning of focus groups* [1], *think-aloud protocol* [7, 49], *questionnaires* [11], *understanding of metaphors and interface design* [11, 13], *non-verbal cues* [47], *web design* [24], *objective and subjective*

measures of usability [19], and *structured interviews* [42]. Currently, the CULTUSAB project [8] is conducting a systematic empirical research program of investigating cultural aspects of the practice of usability in general and usability evaluation methods in particular [e.g., 6, 7-10].

1.1 Culture, Collaboration, and Usability

Existing research in cultural usability research has largely focused on aesthetic issues, methodological aspects, and practitioner concerns. There is little empirical research on the cultural aspects of usability in computer supported collaboration environments. In this paper, we present usability analysis of an experimental study of intra- and inter- cultural computer supported collaborative learning of Chinese and Anglo-American participants.

1.2 Computer Supported Intercultural Collaboration

Computer supported intercultural collaboration (CSIC) is an emerging field of study centrally concerned with the iterative design, development, and evaluation of technologies that enhance and enrich effective intercultural communication and collaboration. There are two interrelated aspects of interaction design in developing CSIC systems: (i) interacting with computers and (ii) interacting with other persons. Both these aspects of interaction can be influenced strongly by culture, given the strong empirical evidence documenting cultural differences in *cognition* [26], *communication* [17], *behavior* [21], and *interacting with computers* [43]. In line with that research program articulated in [41], the research project discussed here originally focused on the cultural influence on (i) how participants from similar and different cultural backgrounds appropriate affordances in a CSIC environment [40, 41] and (ii) how participants relate to each other during and after computer supported collaborative interaction [40, 41, 44]. This paper focuses on usability aspects of the research project. The analytical objective is to investigate cultural aspects of usability in computer supported collaboration in general. Before discussing the methodological aspects of the research project, the key definitions of socio-technical affordance, appropriation of affordance, and technological intersubjectivity are provided below.

1.3 Definition of Socio-Technical Affordance

In computer-supported collaboration, each actor is both a user of the system as well as a resource for the other users. *Technology affordances* are action taking possibilities and meaning making opportunities in a user-technology system. Similarly, *social affordances* are action taking possibilities and meaning making opportunities in a social system. In socio-technical systems that

facilitate collaboration, these amalgamate into *socio-technical affordances*. There has been little work on the concept of socio-technical affordances. Drawing upon foundational work in ecological psychology on the formal definition of affordances [34, 39], the following definition is offered for *socio-technical affordance.*

Let Wpqr (e.g., person-sending-email-to-another-person system) = (Tp, Sq, Or) be composed of different things T (e.g., email technology); S (e.g., email sender) and O (e.g., email receiver). Let p be a property of T; q be a property of S and r be a property of O. The relation between p, q and r, p/q/r, defines a higher order property (i.e., a property of the socio-technical system), α. Then α is said to be a socio-technical affordance of Wpqr if and only if:

(i)   Wpqr = (Tp ,Sq, Or) possesses α

(ii) Neither T ,S, O, (T, S), (T,O), (S,O) possesses α

The formal definition of socio-technical affordance provided above reflects the duality of individuals' perception with respect to the technology as well as other persons. The formal definition informs the design of experimental studies of computer supported intra- and inter-cultural collaboration. Systematic variation of each of the three elements—technology (T), self (S), and other (O)—generates studies of the appropriation of affordances as higher-order relational properties.

1.4 Appropriation of Affordances
The concept of affordance simultaneously specifies the two concurrent levels of meaning and action. Although the perception of objects and other persons can be accounted by a Gibsonian ecological approach, the perception of events cannot be accounted on strictly ecological grounds [33]. Gibson's rejection of a role for higher order cognitive processes in perception-action is problematic, as the perception of events has interactional consequences in computer-supported collaboration. Interactions in socio-technical environments are a dynamic interplay between ecological information as embodied in artifacts and individual actions grounded in cultural schemas. The essential mediation of all interaction is the central insight of socio-cultural theories of the mind [46]. The conceptualization of interaction as being mutually "accountable" (observable and reportable) is the critical insight of ethnomethodology [16]. Following these two schools of thought, interactions in socio-technical systems are conceptualized as appropriation of socio-technical affordances. Adapting Stoffregen's discussion of behavior [34, p.125], appropriation is defined as "*what happens at the conjunction of complementary affordances and intentions or goals*. Research into social aspects of HCI [28] has shown that even computer-literate users tend to use social rules and display social behavior in routine interactions with computers. Social interaction is grounded strongly in culture as every person carries within patterns of thinking, feeling, behaving and potential interacting. Thus, participants in computer supported collaboration make culturally appropriate and sensitive choices and decisions in their actual appropriation of affordances.

1.5 Technological Intersubjectivity
Information and Commutation Technologies (ICT) and the Internet have changed our social relations with others and objects in fundamental ways that sometimes transcend their very technology mediation. Our interactions with others and objects

are increasingly informed by the operational logic of technology, hence technological intersubjectivity. For example, technology lets us assign distinct ring tones, images, priorities to our significant others. Our psychological perception and phenomenal relation with social others is being increasingly transformed by the advances in information and communication technologies and social software. Human beings are not only functional communicators but also hermeneutic actors. In technological intersubjectivity, technological mediation can sometimes (but not necessarily always) disappear like in Clarke's third law of technology [5]. Technological intersubjectivity is an emergent resulting from psychological-phenomenological nexus of the electronic self–other social relationship.

1.5.1   Definition of Technological Intersubjectivity
Technological intersubjectivity (TI) refers to a technology supported interactional relationship between two or more actors. TI emerges from a dynamic interplay between the technological relationship of actors with artifacts and their social relationship with other actors.

From a functional perspective, psychological intersubjectivity doesn't require two or more persons to have the same or similar subjective experience. Put differently, having a collective phenomenal experience is not a necessary condition for psychological intersubjectivity. In psychological intersubjectivity, the other human being is always an object of our attention and an object in our awareness. We observe the other person for communicative cues and informational structures relevant to the ongoing interaction. Unlike in phenomenological intersubjectivity there is no requirement for an emphatic relationship with the other person. However, in the emergent technological case, there is a dynamic interplay between these psychological and phenomenological aspects. In technological intersubjectivity, information processing entailed by computational support can enhance and enrich the communicative possibilities and communion potentials of two or more human beings. Socio-technical systems and online communities have potentials for both psychological and phenomenological intersubjective experiences without the requirement that interacting persons be co-present in the same place and interact at the same time.

2.   METHODOLOGY
The formal definition of socio-technical affordance a in Wpqr = (Tp, Sq, Or) has two important elements: technology T and individual actors S, O [40, 41, 44]. The definition of appropriation of affordances has two important elements: affordances and intentions[41]. Based on these two definitions, an experimental study was designed that introduced a variation in the cultural background of individuals (by selecting participants from a nation-state based ethnically stratified random sampling frame) but kept invariant the technological interface T and interactional setting. Briefly, the experimental study investigated how pairs of participants from similar and different cultures (American-American, American-Chinese, and Chinese-Chinese) appropriated affordances in a quasi-asynchronous computer supported collaborative learning environment with external representations in order to collaboratively solve a public health science problem. Usability analysis reported in this paper was conducted on the empirical data generated by this experimental study.

2.1 Research Questions

Four separate lines of empirical research have demonstrated that culture influences: **(1)** *social behavior,* [21], **(2)** *communication* [17], **(3)** *cognitive processes* [26], and **(4)** *interacting with computers* [43]. These four lines of empirical research were integrated into a methodological framework (Figure 1).
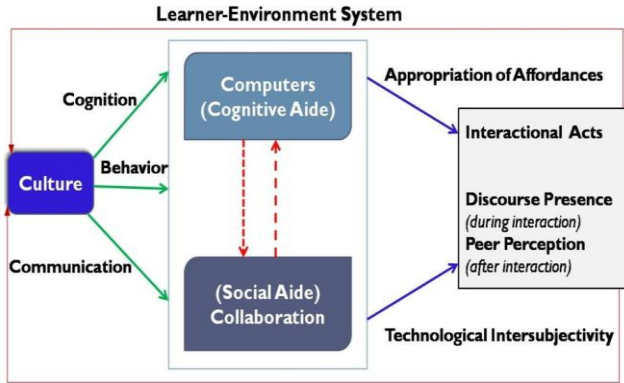

Figure 1: Methodological Framework

The primary purpose of the study was to answer two basic research questions [see 40, 41]. The first research question asked "*to what extent does culture influence the appropriation of socio-technical affordances*? The second research question asked *"to what extent does culture influence technological intersubjectivity?"*

An experimental study was originally designed and conducted to answer these two research questions. Several theoretical predictions were generated from prior empirical evidence [17, 21, 26] warranting the claim that both the perception and appropriation of affordances might vary across cultures and that interpersonal perceptions and relations will vary across cultures [see 40, 41]. As mentioned earlier, this paper presents usability analysis of the empirical data generated by the experimental study. Specifically this paper seeks to answer the following two research questions on cultural usability in computer supported collaboration settings:

1. To what extent does culture influence *objective usability measures of usability* in computer supported collaboration?

2. To what extent does culture influence *subjective usability measures of usability* in computer supported collaboration?

In the next five subsections experimental design, materials, research hypotheses, sampling and procedure are discussed briefly.

2.2 Experimental Design

The experimental study design consisted of three independent groups of dyads from similar and different cultures (Anglo-American, Chinese) doing collaborative problem solving in a knowledge mapping learning environment. The three experimental conditions were the Chinese-Chinese *intra*-cultural condition, the American-American, *intra*-cultural condition and the Chinese-American *inter*-cultural condition.

In all three experimental conditions, the collaborative dyads were given the same experimental task. All the collaborative

dyads interacted in the same computer supported collaborative learning environment after reading the same instructions, software tutorial and demonstration. The same instruments were administered to all participants. Internal validity and external validity were actively considered when designing and conducting the experiment. Construct validity was addressed by using existing instruments with high validity and reliability [2, 21, 31, 35]. Brief descriptions of the experimental study's software and topics are provided next.

2.3 Materials

2.3.1 Software

The computer supported collaborative learning environment used in this experimental study has an "information viewer" on the left in which materials relevant to the problem are displayed. This information viewer functions as a simple web browser, but the presentation of materials is constrained as discussed in the next section. The environment has a shared workspace or "information organizer" on the right hand side in which participants can share and organize information they gather from the problem materials as well as their own interpretations and other ideas. The "discussion" tool below the "information viewer" on the left enables participants to discuss their ideas in a threaded discussion format. Figure 2 below displays a captioned screenshot of the environment used in the experimental study.
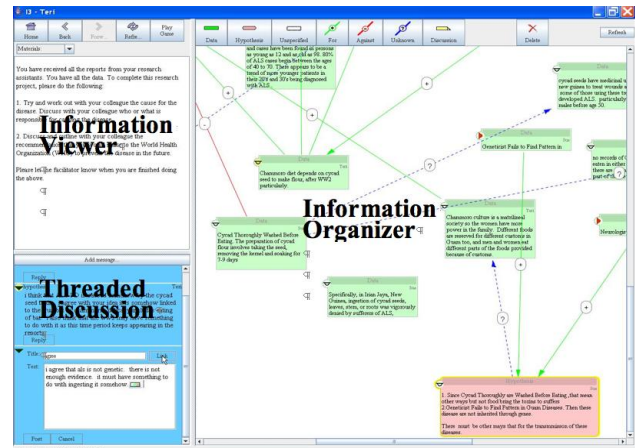

Figure 2: Screenshot from I3P1's session

The "information organizer" workspace includes tools derived from Belvedere [36] for constructing knowledge objects under a simple typology relevant to the experimental task of identifying the cause of a phenomenon (e.g, a disease), including data (green rectangles, for empirical information) and hypotheses (pink rectangles, for postulated causes or other ideas). There are also linking tools for constructing consistency ("for") and inconsistency ("against") relations between other objects, visualized as green links labeled "+" and red links labeled "-" respectively. "Unspecified" objects and "unknown" links are also provided for flexibility. Finally, an embedded note object supports a simple linear (unthreaded) discussion that appears similar to a chat tool; except that a note is interactionally asynchronous and one can embed multiple notes in the knowledge map and link them like any other object. In the "threaded discussion" section of the environment (see bottom-left of Figure 2 above) participants can embed references to knowledge map objects in the threaded discussion messages by

selecting the relevant one or more graph object while composing the message. The references show up as small icons in the message. When the reader selects the icon, the corresponding object in the knowledge map is highlighted, indicating the intended referent.

Mutual awareness of participants' *artifacts* is supported in the software environment as follows: all knowledge map nodes and threaded discussion messages carry the name of the participant who first created it. The mutual awareness features of *artifacts* and of *activity* are shown in Figure 2, a screenshot taken from I3P1 (I stands for Anglo-American--Chinese inter- cultural session, 3 stands for the number of experimental session in this condition, P1stands for Participant 1). In Figure 2, the I3P1's screen name of "Teri" (screen name selected by participant) appears on the title bar of the application window and on knowledge map nodes and message created by I3P1. Similarly, I3P2's screen name of "Sue" appears on artifacts created by him. Artifacts marked with a solid red triangle in the top right corner are from I3P2 and are yet to be opened by I3P1. The yellow circle on the threaded discussion message of I3P1 in the lower left region of Figure 2 indicates artifacts created by "Teri" (I3P1) but not yet read by the study partner, "Sue" (I3P2). Thus each participant is aware of the new artifacts from the study partner as well as the artifacts not yet read by their study partner. In the lower-left corner of Figure 2 shows I3P1 appropriating the affordances for referencing knowledge map artifact (yellow outlined hypothesis node in the bottom-right of Figure 2).

2.3.2   Protocol for Workspace Updates
To simulate asynchronous online interactions, the actions of each participant in the shared workspace were not displayed immediately in the other participant's workspace. As a person worked, the actions of that person were sent to the other participant's client application, but were queued rather than displayed. Participants were given a new report after playing the game of Tetris™. Tetris™ was chosen as it presents a different sensory-motor perceptual task than the primary experimental study task of collaborative knowledge map co-construction and simulates taking a break from the studies in real-worlds asynchronous learning settings [36]. After the game of Tetris™, all of the currently queued actions on that client were displayed. Conflicts that might arise when both participants edited the same object were resolved through operational transformations [36]. The delayed updating protocol simulates one aspect of the experience of asynchronous collaboration: a participant sees what one's partner has done upon returning to a workspace after a period of time. It excludes the possibility of synchronous conversation in which one participant posts a message in the workspace and receives an immediate reply. The "refresh" feature of the software enables one to get all updates to that point in time.

2.3.3   Alternates for Action
The software environment provides multiple alternatives for appropriation of affordances. For example, participants can discuss with each other using the threaded discussion tool or the embedded notes tool. Participants can also use the knowledge-map objects to discuss the task at hand or any other topic of interest. Participants can refer to artifacts by deictic referencing (this, that, etc…) or use the cross-referencing feature of the threaded discussion. Participants can externalize the perceived relations between their concepts by creating external evidential relations between objects in the knowledge-map, by spatial arrangement, or by mentioning them in discussion. Participants have multiple ways of sharing the information presented to them (threaded discussion, embedded notes, and knowledge-map).

The research strategy was to provide participants with a feature rich collaborative environment with multiple alternates for action. By incorporating systematic variation in the assignment of participants to the collaborative dyad based on their cultural background and gender, the experimental design measured and observed systemic differences in how participants used the tools and resources of the technology (research question 1, appropriation of affordances) and related to each other during and after their interaction (research question 2, technological intersubjectivity).

2.3.4   Topics
The study presented participants with a "science challenge" problem that requires participants to identify the cause of a disease known as ALS-PD on the island of Guam. This disease has been under investigation for over 60 years, in part because it shares symptoms with Alzheimer's and Parkinson's diseases. Only recently have investigators converged on both a plausible disease agent (a neurotoxic amino acid in the seed of the Cycad tree) and the vector for introduction of that agent into people (native Guamians' consumption of fruit bats that eat the seed). Over the years numerous diverse hypotheses have been proposed and an even greater diversity of evidence of varying types and quality explored. These facts along with the relative obscurity, multiple plausible hypotheses, contradicting information, ambiguous data and high interpretation make this a good experimental study task for measuring cultural effects on appropriation of affordances and on technological intersubjectivity.

All experimental study materials were in English. All participants began with a mission statement that provided the problem description and task information. Four mission statements corresponding to the four participant assignment configurations (Chinese vs. Anglo-American x P1 vs. P2) were administered (http://lilt.ics.Hawai'i .edu/culturalreps/materials/). Due to the distribution of conflicting evidence, sharing of information across participants and study sessions is needed to expose the weakness of genetics as well as to construct the more complex explanation involving bats and cycad seeds. Given the nature of the information distribution between the two participants, working out the consumption of bats as an optimal hypothesis involves making these cross-report collaborative connections and also considering and rejecting other probable factors. The study task and task materials are designed to highlight "social division of cognitive labor". The experimental study encouraged participants to interact with each other by including the following reinforcing task instruction on each report (set of 4 articles): "*Please share and discuss this information with you colleague. Please play the game to receive the next report from your research assistant.*" The next section discusses several research hypotheses generated from the culture theory and empirical findings in cross-cultural psychology.

2.4 Participants
Participants were recruited from the graduate student community at the University of Hawai'i at Mānoa. Each participant was offered a payment of US$75 for participating in

the study. Participant selection and treatment assignment are discussed next.

2.4.1 Sampling
There is a tendency in cross-cultural computer mediated communication research to use cultural models bounded by modern nation-states. Although nationality based stratified sampling frames remain a methodologically convenient way to select participants provided. cultural homogeneity of the participants is not to be assumed but empirically measured. We used the PVQ individual values survey [31] and the GLOBE [21] instrument to empirically assess differences in the two participant groups at the individual and group levels respectively.

2.4.2 Assignment
Participants were randomly assigned to either the intra- or the inter- cultural profiles and the same or different gender profiles. Excluding 6 pilot studies, a total of 33 experimental sessions involving 66 pairs of participants were conducted. Data from 3 experimental sessions was discarded due to issues of a missing screen recording, a software crash, and a disqualification. There were 10 pairs of participants for each of the three treatment groups: Chinese-Chinese intracultural; Anglo-American-Anglo-American intracultural, and Anglo-American-Chinese intercultural groups. All the three conditions were gender-balanced because gender can substantially influence social interaction [37]. Each treatment group included 3 female-female, 3 male-male and 4 female-male dyads.

2.5 Instruments

2.5.1 Demographic Questionnaire
A demographic questionnaire [41, pp.275-276] was administered to collect participants' familiarity with each other, with online learning environments, with usability evaluation studies as well as data about age, gender, ethnic background, duration of stay in the USA , duration of stay in the state of Hawai'i . All participants were requested to make a self-report of their CGPA and also assign a release form for obtaining official records of their CGPA, graduate record examination scores (GRE), and test of English as a foreign language (TOEFL) (Chinese participants only).

2.5.2 Self-Perception: Portrait Value Questionnaire (PVQ)
The 40 item version of the PVQ instrument [41, pp.277-279] recommended for intercultural contexts (Schwartz, S. H, personal communication) was used in the study. The PVQ scale measured cultural values at the individual level. Cronbach's "alpha measures of internal consistency range from .37 (tradition) to .79 (hedonism) for the PVQ (median, .55)" [31, p.532]. Gender specific versions of the self perception PVQ scale were administered.

2.5.3 GLOBE Cultural Dimensions Instrument
The GLOBE instrument [21] was used to measure cultural values at the group level [41, pp.280-293]. Section 1 and Section 3 of the original GLOBE instrument were used in this study. Section 1 of the GLOBE instrument measures a responder's perceptions of their society "Section 1 — The way things are in your society". Section 3 of the GLOBE instrument measures a responder's preferences for their society "Section 3 — The way

things generally should be in your society". According to the "Guidelines for the Use of GLOBE Culture and Leadership Scales," "the construct validity of the culture scales was confirmed by examining the correlations between the GLOBE scales with independent sources (e.g., Hofstede's culture dimensions, Schwartz's value scales, World Values Survey, and unobtrusive measures)" [21]. Phrasing of "this country" has been changed to "my home society" to remove possible ambiguity for Chinese graduate students who might rate Hawai'i , USA instead of the society they grew up in.

2.5.4 Individual Essays
At the end of their collaborative science problem solving, the immediate post-test consisted of each participant individually writing an essay. Identical essay writing instructions were provided to all participants. The instructions asked the participants to (a) state the hypotheses they considered, (b) whether and how their hypotheses differed from those of their study partners', and (c) their final conclusion.

2.5.5 Peer-Perception: Portrait Value Questionnaire (PVQ)
Technological intersubjectivity after interaction was measured by the second immediate post-investigative-test. This was the administration of the Portrait Value Questionnaire (PVQ) [31] instrument with a reversal of the direction of assessment [41, pp.304-306]. This time instead of assessing themselves, participants assessed their collaborative partners. Based on their collaborative interactions, each participant rated his/her impressions of the study partner on the Portrait Value Questionnaire (PVQ).

2.5.6 Acculturation: SL-ASIA Questionnaire
Acculturation is a process that occurs when two or more cultures interact together. This becomes an external variable in cross-cultural research conducted with participants from an immigrant culture in a host culture (in our case, Chinese participants in Hawai'i , USA. Although one could argue that Hawai'i is different culture for Anglo-American participants who grew up on mainland USA). This external variable can be controlled by measuring the acculturation level of the participants belonging to the minority immigrant culture [38]. Participants with high level of acculturation can be best used as members of the majority host culture or not included in the study [38]. This research project used the Suinn-Lew Asian Self Identity Acculturation (SL- ASIA) scale [35] to measure the acculturation levels of the Chinese participants [41, pp.307-311]. This scale was chosen as it is specifically designed for Asians. Suinn et al. [35] reported an internal-consistency estimate of .91 for the SL-ASIA instrument.

2.5.7 Intercultural Sensitivity: Intercultural Sensitivity Instrument
Intercultural sensitivity is a vital skill for intercultural collaborations [2]. The SL-ASIA scale provided a measure of Chinese participants' assimilation to USA. The intercultural sensitivity instrument (ICSI) [2] was used to measure Anglo-American participants' self-assessment of intercultural sensitivity [41, pp.312-315]. Bhawuk and Brislin (1992) report that "the ICSI was validated in conjunction with intercultural experts at the East-West Center with an international sample (n=93)" (p. 423). The word "Japan" in the original ICSI scale

was changed to "China" to fit the context of Chinese-American collaboration setting of the experiment. Part three of the original ICSI instrument was not used, as pilot studies indicated that it was irrelevant to the purposes of this experimental study.

2.5.8 User Satisfaction: QUIS Questionnaire
The QUIS 7.0 questionnaire [18] was administered to collect the participants subjective perceptions and preferences of the learning environment [41, pp.316-321]. The QUIS has high reliability (Cronbach's alpha = 0.95 and high construct validity (alpha = 0.86) [18].

2.6 Procedure
Two students participated in each session. Experimental sessions lasted about 3.5 hours on average. Informed consent was obtained from all participants for both the pilot studies and the experimental studies. After signing the informed consent forms, participants completed a demographic survey. They were then given CGPA/GRE/TOEFL score release form, Self-Perception PVQ [31], and the GLOBE instrument [21]. After completing these three forms, participants were brought into a common room. Participants were then introduced to the software and the structure of the experimental study through an identical set of instructions and demonstrations across all three conditions.

After the software demonstration, the two participants were led back to their respective workstations in two different rooms. They were then instructed to begin work on the study task. Participants had up to 90 minutes to work on the information available for this problem. The update protocol described in [36] was used to synchronize the workspaces of the two participants. At the conclusion of the investigative session, each participant was given up to 30 minutes to write an individual essay. The CSIC environment remained available to each participant during the essay writing, but the participants were requested not to engage in any further communication. After each participant had finished writing the individual essay, the other-perception PVQ instrument [31] and the QUIS instrument [18] were administered. This concluded the experimental session. Participants then completed the payment forms and were debriefed.

3. RESULTS
Results are grouped under the five subsections of *demographics*, *culture measures*, *objective usability measures*, and *subjective usability measures*. The empirical data generated by the experimental study were analyzed at four levels: culture (Anglo-American, Chinese), gender (female, male), dyadic culture (American-American, American-Chinese, Chinese-Chinese), and dyadic gender (female-female, female-male, male-female). Unless otherwise indicated, the statistical summaries refer to two-way analysis of variance with respect to culture (Anglo-American, Chinese) and gender (female, male).

3.1 Demographics
There was no age difference at any of the four levels of analysis (culture, gender, dyadic culture, dyadic gender). As expected, Anglo-American participants reported to have spent significantly more time in the United States of America than the Chinese participants. On the other hand, the time spent by the participants in Hawai'i with respect to culture and gender was not statistically significant. There were no significant differences at any of the four levels of analysis for prior experience with experimental studies, prior knowledge about the experimental task, and partner familiarity.

3.2 Culture Measures
As mentioned before, Portrait Values Questionnaire (PVQ) [31] was used to measure culture at the individual level. The GLOBE instrument [21] was used to measure culture at the group level.

Ten individual values are measured by the Portrait Values Questionnaire (PVQ) [31]. Ten individual values are measured by the Portrait Values Questionnaire (PVQ) [31]. Statistical analysis showed that at the level of culture, the PVQ values of Conformity($F(1,56)=7.71,p=0.008$),Benevolence($F(1,56)=5.60, p=0.02$),Universalism($F(1,56)=6.66,p=0.01$), Self-Direction($F(1,56)=7.48,p=0.01$),Stimulation($F(1,56)=10.02,p=0.003$) and Security.

Significant differences were observed on both sections of the GLOBE instrument. For the "AS IS" section, significant differences between the American and Chinese groups were observed for Institutional Collectivism ($F(1,56)=43.55, p<0.01$), In-Group Collectivism ($F(1,56) =102.43,p<0.01$), and Assertiveness ($F(1,56)=28.57, p<0.01$). For the "SHOULD BE" section of the GLOBE instrument, statistically significant differences were found for Uncertainty Avoidance ($F(1,56)=49.65,p<0.01$), Assertiveness ($F(1,56)=4.20,p=0.04$), Future Orientation ($F(1,56)=14. 23, p=0.01$),Humane Orientation ($F(1,56)= 7.90,p=0.007$),and Gender Egalitarianism ($F(1,56)=4.89, p=0.03$).

In summary, there is necessary and sufficient evidence to conclude that Chinese and Anglo-American participants significantly differ on specific PVQ individual values as well as GLOBE cultural dimensions. Even though a nation state based stratified random sampling frame was utilized; systemic variation between the two participant groups is thus empirically documented and not stereotypically assumed or dogmatically asserted.

3.3 Objective Usability Measures
Objective usability measures consisted of the *efficiency* (total task time in minutes), and *effectiveness* (usage of certain features of interest). Each of these is discussed below.

3.3.1 Efficiency
On average, task time was greater for Chinese participants (M=156.07 minutes, SD=19.22) than the Anglo-American participants (M=144.96, SD=25.14). On average, female participants' task time (M=155.58, SD=20.88) was greater than the male participants (M=145.44, SD=24.00) in the study. A two-way ANOVA showed marginal main effects for culture ($F(1,56)$= 3.77, p=0.06) and gender ($F(1,56)$=3.14, p=0.08). On the other hand, total task time varied significantly between the intra- and inter-cultural conditions of the experimental study ($F(2,51)$=5.17, p=0.009). A Bonferroni post-hoc comparison showed that the Anglo-American intra-cultural group had significantly lower task time than the Chinese intra-cultural group and the American-Chinese inter-cultural group. No significant differences were observed at the dyadic gender level.

3.3.2 Effectiveness
Effectiveness measures consisted of the software features of structural and functional significance to computer supported

collaborative learning. Each measure is introduced, briefly discussed, and then empirical results are presented.

3.3.2.1 Cross-referencing
Video analysis of the screen recordings of participant sessions was done to obtain the counts for the referencing. No statistically significant differences were found at any of the four levels of analysis (culture, gender, dyadic culture, dyadic gender).

3.3.2.2 Shared workspace refresh
As discussed under software in the methodology section, the shared workspace (information organizer +discussion) could be refreshed (a) automatically after returning from game or (b) on demand when the participant clicks on the "Refresh" button (see top-right in Figure 2). Recall that there were four reports and a final page. Participants had to play and quit the game in order to receive the next report. All the participants played and quit the game at least four times and therefore, received all the four reports. However, the refresh count varied due to the differences in the number of on-demand refreshes of the shared work-space. There was no significant main effect for refresh count with respect to culture. However, the refresh count was lower for the female participants with a marginal main effect for gender ($F(1,56)=3.50, p=0.067$). At the dyadic gender level of analysis, a marginally significant effect was found ($F(2,51)=2.632, p=0.82$). A post-hoc Bonferroni comparison showed a marginally significant difference between the male-male and female-female collaborative dyads.

3.3.2.3 Threaded discussion messages
Counts for *discourse usage* were obtained from the software logs of participant sessions. For threaded discussion messages, Anglo-American participants created more threaded discussion messages than the Chinese participants and the difference was statistically significant ($F(1,56)=8.88, p=0.004$).

3.3.2.4 Embedded discussion notes
For the embedded discussion notes, no statistically significant difference was found. However, the observed empirical trend was that Chinese participants created more embedded discussion notes than the Anglo-American participants.

3.3.2.4 Evidential relation links
Counts for *evidential relation links* were obtained from the software logs of participant sessions The observed empirical evidence was exactly opposite to that of the theoretical prediction at the level of culture ($F(1,56)=5.54, p=0.02$) for the number of evidential relation links created. On average, Anglo-American participants created more evidential relation links compared to the Chinese participants of the experimental study.

3.3.2.5 Knowledge-map nodes
No significant differences were observed between the Chinese and Anglo-American participants in the number of "data" and "hypotheses" nodes created. However, Chinese participants created less number of "unspecified" nodes than the Anglo-American participants ($F(1, 56)=5.76, p=0.02$). At the gender level of analysis, female participants created significantly more "hypothesis" nodes than the male participants ($F(1, 56)=4.68, p=0.035$).

3.3.2.6 Verbosity
Language remains a potential mediating variable in the performance of Chinese participants. To empirically evaluate this mediating variable, session verbosity (total words individually produced by a participant in the collaborative session) and essay verbosity (total words produced by a participant in the individually written essay) were calculated. A two way analysis of variance for session verbosity showed significant main effects for both culture ($F(1,56)=4.46, p=0.39$) and gender ($F(1,56)=6.70, p=0.012$). On average, Anglo-American participants produced more words in the collaborative session than the Chinese participants. Female participants produced more words in the collaborative session than the male participants. Similar results were obtained for the essay verbosity measure.

3.4 Subjective Usability Measures
As mentioned earlier, the validated usability instrument, QUIS questionnaire [4] was administered to collect the participants subjective perceptions and preferences of the learning environment. The QUIS 7.0 instrument also measured participants' subjective satisfaction with the instructions and the software tutorial besides various systems measures. The coding key for the QUIS instrument was used for the quantitative analysis of the data (http://lap.umd.edu/QUIS/QuantQUIS.htm). Table 2 below presents a summary of the QUIS results with respect to culture.

Table 1: Summary of QUIS Results with respect to Culture

QUIS Construct	Anglo-American Mean (SD)	Chinese Mean (SD)	ANOVA F-Statistic (p-value)
Overall Reaction	6.18(1.35)	6.49(1.22)	0.84(0.36)
Screen	6.95(1.06)	6.15(1.19)	8.00(0.01)*
Terminology &System Information	7.46(1.02)	6.81(1.22)	4.84(0.03)*
Learning	7.78(1.09)	7.22(1.32)	3.17(0.08)
System Capabilities	7.17(1.13)	6.85(1.43)	0.92(0.34)
Tutorial	7.90(1.03)	7.71(1.03)	0.49(0.48)

* $p<0.05$

3.4.1 Overall user satisfaction
On average, the overall user reaction for the Chinese participants was higher than Anglo-American (see table 1). However, no significant differences were found at any of the four levels of analysis (culture, gender, dyadic culture, dyadic gender).

3.4.2 Perception of information display
Significant differences were observed between the Chinese and Anglo-American participants on the QUIS section for information display on the screen (see table 1). Chinese participants' subjective satisfaction scores for the screen information display were lower than the Anglo-American participants.

3.4.3 Perception of system terminology

System Terminology and System Information section of the QUIS instrument received significantly lower ratings from the Chinese participants.

3.4.4 Perception of system capabilities

No significant differences were observed despite lower scores by Chinese participants compared to the Anglo-American participants of the experimental study.

3.4.5 Perception of ease of learning

Results for the Learning section of the QUIS instrument showed a marginally significant difference on the ease of learning measure at the level of culture.

3.4.6 Perception of tutorial

Results for the Tutorial section of the QUIS instrument showed no significant difference for participants' subjective evaluation of the software demo and experimental instructions at any of the four levels of analysis. Therefore, experimenter bias and "demand characteristics" [27] are ruled out as confounding variables in the study.

In summary, there is a discrepancy between Chinese participants' higher overall satisfaction ratings and their lower satisfaction ratings for the specific components of the system (**Screen** and **Terminology & System Information).** Similarly, Anglo-American participants' reported lower overall reaction ratings but higher satisfaction ratings to the specific components of the system.

3.4.7 Analysis of comments

The QUIS instrument includes an open-ended comments solicitation at the end of each of the six sections. The user comments were transcribed. A few illustrative user comments are included below:

C2P2: "The link function is very helpful and I'll expect a drag and drop from the organizer to the message panel."

C6P2: "Having a zoom in/out feature might help."

I3P1: "I like the idea and that can link data boxes. However, a function that would put everything into a condensed list or a short outline to see all data at once should be helpful b/c sometimes too much information is displayed at once to work with in a rational manner."

I8P2: "In general the system speed is satisfactory, and it's reliable, but it needs more on other functions such as undo, correcting typo, etc.

I11P2: "It is good if there is a "undo" and "redo" (ctrl+z or ctrl+y)."

A3P1: "When new text boxes appear from the partner, they should be in a separate section so it is easy to see them and sort them out from mine and older ones. They should be color coded differently until read. The size of screens should be adjustable to allow more."

A7P2: "Instructions were well laid out and easy to use. Messages sometimes appear overlapping, difficult to see everything that way."

A8P1: "Overall fairly clear & easy to navigate."

A9P2: "In the boxes, the word "text" should be eliminated in a click. It shouldn't need deleting."

Qualitative analysis of the comments shows that Undo, Copy+Paste, Zooming and Color Coding of Contributions are the most frequent usability suggestions. Usability problems mentioned included scrolling issues. Negative comments were mainly about screen clutter.

The coding scheme developed in [42] was modified and used for the content analysis of the comments. The modified coding scheme is described below:

- **Usability Problem (U):** interaction design flaw or a user difficulty that is directly associated with an interface/interaction design flaw.

- **Suggestion (S):** subjective preference of the participant to the implemented design choice or tradeoff.

- **Positive Comment (P):** participant's subjective approval of a design choice or tradeoff.

- **Negative Comment (N):** participant's subjective disapproval of a design choice or tradeoff.

- **Other Comment (O):** User comment that couldn't be categorized under one of the above categories.

Total Comments = Usability Problems (U) + Suggestions (S) + Negative Comments (N) + Positive Comments (P). Figure 3 presents aggregate comment counts with respect to culture.
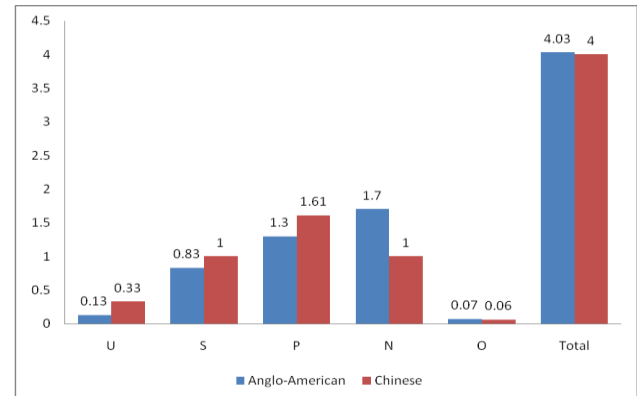

Figure 3: Aggregate Comments across Cultures

Even though Chinese participants made more usability suggestions, more positive comment, and less negative comments than the Anglo-American participants, no significant differences were observed at any of the four levels of analysis.

4. DISCUSSION

The empirical findings reported above can be summarized as follows: Culturally different participants were found to be *engaged* in different collaborative activities and *created* different quantities of artifacts. Further, participants from the two cultural groups had differential user interface *satisfaction* scores along with different quality and quantity of *comments*. The empirical findings discussed add to the literature on the cultural effects on usability evaluation process [e.g., 3, 7, 8, 11-14, 19, 25, 32, 47-49]. Specifically, Chinese participants reported higher overall user interface satisfaction scores but

gave significantly lower ratings for information display and terminology aspects of the system. One candidate explanation for this empirical finding is the cultural-cognitive difference in holistic vs. analytical reasoning [26]. Participants with an analytical thinking style might offer more specific user interface satisfaction ratings of the individual components of the system compared to participants with a holistic thinking style. This could account for the greater specificity of Anglo-American participants with individual aspects of the socio-technical system compared to Chinese participants' overall assessment of the socio-technical system. Cultural concerns with "face saving might be another explanation.

4.1 Implications for Research

Honold [20] identified eight factors for cultural usability: *objectives of the users*; *characteristics of the users*; *environment*; *infrastructure*; *division of labor*; *organization of work*; and *mental modes based on previous experience and tools*. *Objectives of users* have been the focus of recent work in cultural usability that focuses on evaluator-subject interpersonal relations [32, 42]. Cultural usability should consider the debate around "demand characteristics" [27] in experimental psychology. A better understanding of demand characteristics is of benefit to the field of usability in general. *Characteristics of user* are often ignored the concept of culture is rarely operationalized. Cultural characteristics are attributed to individuals by virtue of ethnic affiliation or nation-state membership. Given the "fading quality of culture" the homogenous assumptions of culture are unwarranted [29]. As stated above under culture measures in the results section, cultural characteristics of participants in this experimental study are assessed, evaluated, and documented at level of individual values as well at the level of group dimensions. In computer supported collaboration, *environment*; *infrastructure*; *division of labor*; *organization of work*; and *mental modes based on previous experience and tools* all become crucial factors of concern. For example, despite the differences in cultural characteristics of the *users*, their *activities*, their *artifacts*, and their *subjective interface perceptions*, as reported in [see 40, 41], no significant differences were found in individual learning outcomes. In a prior contribution, we have proposed [45] a comprehensive design evaluation framework of usability, sociability, and learnability for computer supported collaborative learning (CSCL) systems.

In my opinion, cultural usability needs to move beyond the comparative method approach to usability evaluation. Computer supported collaboration in intra- and inter- cultural contexts presents some unique challenges which do not entirely fall under the purview of usability typically conceived of as efficiency, effectiveness and satisfaction. Further, in the established HCI research field of computer supported cooperative work (CSCW), the emerging HCI field of human-information interaction (HII) [22], the technology enhanced learning paradigm of computer supported collaborative learning (CSCL), and the emerging field of computer supported intercultural collaboration (CSIC), users are interacting with each other as well as with the computer. Currently, the Internet is undergoing a profound shift towards a participatory mode of interaction. With the advent of fundamentally social software such as social networking sites (Orkut, Facebook, MySpace, Mixi etc), cultural usability needs to be expanded to include technological intersubjectivity [41]. An initial attempt to address theoretical and methodological issues on intra- and inter- cultural computer supported collaboration can be found in [41]. Specifically, [41] attempted to develop a theory-based empirical study of socio-technical affordances and their appropriation and an empirically informed theory of the structures and functions of technological intersubjectivity. In my opinion, there is a need for a real-time and real-space interactional account of cultural cognition for the field of cultural usability. In sum, this is a call for the field of cultural usability to widen its research horizons.

5. ACKNOWLEDGEMENTS

6. REFERENCES

1. Beu, A., Honold, P. and Yuan, X. How to Build Up an Infrastructure for Intercultural Usability Engineering. *The International Journal of Human-Computer Interaction*, *12* (3&4). 347-358.

2. Bhawuk, D.P.S. and Brislin, R.W. The measurement of intercultural sensitivity using the concepts of individualism and collectivism. *International Journal of Intercultural Relations*, *16* (4). 413-436.

3. Bourges-Waldegg, P. and Scrivener, S.A.R. Applying and testing an approach to design for culturally diverse user groups. *Interacting with Computers*, *13* (2). 111.

4. Chin, J.P., Diehl, V.A. and Norman, K. Development of an instrument measuring user satisfaction of the human-computer interface. in *SIGCHI '88*, ACM., New York, 1988, 213-218.

5. Clarke, A.C. *Profiles of the future: an inquiry into the limits of the possible*. Harper & Row, 1962.

6. Clemmensen, T. and Goyal, S., Cross cultural usability testing. in *The First All India Human Computer Interaction Conference*, (Bangalore, India 2005).

7. Clemmensen, T., Hertzum, M., Hornbæk, K., Shi, Q. and Yammiyavar, P. Cultural Cognition In The Thinking-Aloud Method For Usability Evaluation *International Conference on Information Systems - ICIS2008*, Paris, Accepted for publication.

8. Clemmensen, T. and Plocher, T. The Cultural Usability (CULTUSAB) Project: Studies of Cultural Models in Psychological Usability Evaluation Methods. *Lecture Notes in Computer Science*, *4559*. 274.

9. Clemmensen, T. and Qingxin, S., What is part of a usability test? in *CHI '08 extended abstracts on Human factors in computing systems.*, (2008).

10. Clemmensen, T., Shi, Q., Kumar, J., Li, H., Sun, X. and Yammiyavar, P. Cultural Usability Tests–how usability tests are not the same all over the world. *Proceedings of Conference of Human Computer Interface, International. Beijing, PR China*.

11. Day, D. and Evers, V., Questionnaire Development for Multicultural Data Collection. in *Proceedings of the International Workshop on Internationalisation of Products and Systems*, (1999).

12. Day, D. and Evers, V. The Role of Culture in Interface Acceptance. *Human Computer Interaction, INTERACT' 97*.

13. Evers, V., Cross-cultural Understanding of Metaphors in Interface Design. in *Attitudes toward Technology and Communication*, (London, 1998).

14. Evers, V., Kukulska-Hulme, A. and Jones, A., Cross-Cultural Understanding of Interface Design: A Cross-Cultural Analysis of Icon Recognition. in *Proceedings of the International Workshop on Internationalization of Products and Systems*, (Rochester, 1999).

15. Fernandes, T. *Global Interface Design*. Academic Press, 1995.

16. Garfinkel, H. *Studies in Ethnomethodology*. Prentice-Hall, Englewood Cliffs, NJ, 1967.

17. Hall, E. *Beyond Culture*. Anchor Books, New York, 1977.

18. Harper, B., Slaughter, L. and Norman, K. Questionnaire administration via the WWW: A validation & reliability study for a user satisfaction questionnaire. www.lap.umd.edu/webnet/paper.html.

19. Herman, L. Towards Effective Usability Evaluation in Asia: Cross-Cultural Differences. *Proceedings of the 6th Australian Conference on Computer-Human Interaction (OZCHI'96)*. 135.

20. Honold, P. Culture and Context: An Empirical Study for the Development of a Framework for the Elicitation of Cultural Influence in Product Usage. *International Journal of Human-Computer Interaction*, *12* (3&4). 327 - 345.

21. House, R.J., Hanges, P.J., Javidan, M., Dorfman, P.W. and Gupta, V. *Culture, Leadership and Organizations: The GLOBE study of 62 societies*. Sage Publications, Newbury Park, CA, 2004.

22. Jones, W., Pirolli, P., Card, S.K., Fidel, R., Gershon, N., Morville, P., Nardi, B. and Russell, D.M. "It's about the information stupid!": why we need a separate field of human-information interaction *CHI '06 extended abstracts on Human factors in computing systems*, ACM, Montreal, Quebec, Canada, 2006, 65-68.

23. Khaslavsky, J. Integrating Culture into Interface Design. *CHI 98 conference summary on Human factors in computing systems*. 365-366.

24. Marcus, A. and Gould, E.M. Cultural Dimensions and Global Web User-Interface Design. *Interactions*, *7* (4). 32 - 46.

25. Molich, R., Ede, M.R., Kaasgaard, K. and Karyukin, B. Comparative usability evaluation. *Behaviour & Information Technology*, *23* (1). 65 - 74.

26. Nisbett, R.E. and Norenzayan, A. Culture and Cognition. in Medin, D.L. ed. *Stevens' Handbook of Experimental Psychology*, 2002, 561–597.

27. Orne, M.T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Prevention & Treatment*, *5* (35).

28. Reeves, B. and Nass, C. *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press New York, NY, USA, 1996.

29. Ross, N. *Culture and Cognition: Implications for Theory and Method*. Sage Publications, 2004.

30. Russo, P. and Boor, S. How Fluent is your Interface? Designing for International Users. *Proceedings of the conference on Human factors in computing systems*. 342-347.

31. Schwartz, S.H., Melech, G., Lehmann, A., Burgess, S., Harris, M. and Owens, V. Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology*, *32* (5). 519–542.

32. Shi, Q. and Clemmensen, T. Relationship Model in Cultural Usability Testing. in Aykin, N. ed. *Usability and Internationalization - Second International Conference on Usability and Internationalization, UI-HCII 2007, Held as Part of HCI International 2007, Beijing, China, July 22-27, 2007, Proceedings, Part I* Springer, Heidelberg, 2007, 422-431.

33. Stoffregen, T.A. Affordances and Events. *Ecological Psychology*, *12* (1). 1-27.

34. Stoffregen, T.A. Affordances as Properties of the Animal-Environment System. *Ecological Psychology*, *15* (2). 115-134.

35. Suinn, R.M., Ahuna, C. and Khoo, G. The Suinn-Lew Asian Self-Identity Acculturation Scale: Concurrent and Factorial Validation. *Educational & Psychological Measurement*, *52* (4). 1041-1046.

36. Suthers, D., Vatrapu, R., Medina, R., Joseph, S. and Dwyer, N. Beyond Threaded Discussion: Representational Guidance in Asynchronous Collaborative Learning Environments. *Computers and Education*, *50* (4). 1103-1127.

37. Tannen, D. *Gender and Discourse*. Oxford University Press, 1996.

38. Triandis, H.C., Kashima, Y., Shimada, E. and Villareal, M. Acculturation Indices as a Means of Confirming Cultural Differences. *International Journal of Psychology*, *21*. 43-70.

39. Turvey, M.T. Affordances and Prospective Control: An Outline of the Ontology. *Ecological Psychology*, *4* (3). 173-187.

40. Vatrapu, R. Cultural Considerations in Computer Supported Collaborative Learning. *Research and Practice in Technology Enhanced Learning*, *3* (2). 159-201.

41. Vatrapu, R. Technological Intersubjectivity and Appropriation of Affordances in Computer Supported Collaboration *Communication and Information Sciences*, University of Hawaii at Manoa, Honolulu: Available at http://lilt.ics.hawaii.edu/~vatrapu/docs/Vatrapu-Dissertation.pdf, 2007, 538.

42. Vatrapu, R. and Pérez-Quiñones, M. Culture and Usability Evaluation: The Effects of Culture in Structured Interviews. *Journal of Usability Studies*, *1* (4). 156-170.

43. Vatrapu, R. and Suthers, D. Culture and Computers: A Review of the Concept of Culture and Implications for Intercultural Collaborative Online Learning. in Ishida, T., Fussell, S.R. and Vossen, P.T.J.M. eds. *Intercultural Collaboration I : Lecture Notes in Computer Science*, Springer-Verlag 2007, 260-275.

44. Vatrapu, R. and Suthers, D. Technological Intersubjectivity in Computer Supported Intercultural Collaboration. in *ACM 2nd International Workshop on Intercultural Collaboration*, ACM Digital Library, in press.

45. Vatrapu, R., Suthers, D. and Medina, R., Usability, Sociability, and Learnability: A CSCL Design Evaluation Framework. in *16th International Conference on Computers in Education (ICCE2008)*, (Taipei, Taiwan, in press), (CD-ROM).

46. Wertsch, J. *Vygotsky and the social formation of mind*. Harvard University Press, Cambridge, MA, USA, 1985.

47. Yammiyavar, P. and Goel, K.M., Emphasis on non-verbal cues for interpreting cognitive processes in protocol analysis. in *Indo-Danish HCI Research Symposium* (Guwahati, India, 2006), The Indian Institute of Technology Guwahati, Guwahati 781039. Assam India.

48. Yeo, W.A. Cultural Effects in Usability Assessment. *Proceedings of CHI 98 conference summary on Human factors in computing systems*. 74-75.

49. Yeo, W.A. Global-Software Development Life Cycle: An Exploratory Study. *Proceedings of the SIG-CHI Conference on Human Factors in Computing Systems*. 04 -11.

Xianghong Sun: Do cultural factors affect thinking in usability tests?

Creating a Standardized Corpus of Multimodal Interactions for Enculturating Conversational Interfaces

Matthias Rehm
Franziska Grüneberg
Multimedia Concepts and
Applications
University of Augsburg,
Germany
rehm@informatik.uni-
augsburg.de

Yukiko Nakano
Afia Akhter Lipi
Yuji Yamaoka
Ubiquitous and Universal
Information Environment
Tokyo University of
Agriculture and
Technology, Japan
nakano@cc.tuat.ac.jp

Hung-Hsuan Huang
Graduate School of
Informatics
Kyoto University, Japan
huang@ii.ist.i.kyoto-u.ac.jp

ABSTRACT

To enculturate conversational interfaces it is indespensable to gain insights in culture specific differences of interaction patterns. Information found in the literature is often very general or more of an anecdotal character. In this paper we present a corpus of culture specific interaction patterns that was collected under standardized conditions in Germany and Japan and thus presents a rich source of data for developing interfaces that are tailored to specific cultures.

Author Keywords

cultural computing, embodied conversational agents, HCI

ACM Classification Keywords

H.1.2 [Models and Principles]: User/Machine Systems – human factors, human information processing; H.5.1 [Information Interfaces and Presentations (e.g. HCI)]: Multimedia Information Systems – animations, evaluation/methodology; H.5.2 [Information Interfaces and Presentations (e.g. HCI)]: User Interfaces – theory and methods

INTRODUCTION

Why is there a need for taking the user's culture into account for human computer interaction? Because the user's cultural upbringing gives her heuristics for interpreting behavior and presentations of others. This influences not only face-to-face encounters but has direct consequences for how information is evaluated that is presented e.g. on a website. [15] gives some interesting examples on different styles of information presentation on websites that are influenced by cultural parameters. [10] reports on a study by [23] about influence tactics in different cultures and shows that people from cultures accepting distinct hierarchies (see below) tend to argue by invoking a higher authority, whereas people from cultures with flatter hierarchies tend to argue more friendly and by reasoning. Hofstede has termed these heuristics for behaving and interpreting behavior *mental programs*.

If we take the evidence from the literature seriously that users from different cultures interact based on such culture dependent heuristics, then it is necessary to acknowledge these differences in the design of interfaces. In this paper we focus on embodied conversational agents, which serve as anthropomorphic communication devices and thus create even more severe expectations regarding their behavior (verbal as well as nonverbal). On the other hand, due to this challenge, embodied conversational agents as an interface metaphor have a great potential to realize culture specific interaction behavior in several fields of human computer interaction:

- Information presentation: By adapting their communication style to the culturally dominant persuasion strategy, agents become more efficient in delivering information or selling a point or a product.

- Entertainment: Endowing characters in games with their own cultural background has two advantages. It makes the game more entertaining by i.) providing coherent behavior modifications based on the cultural background and ii.) it let's characters react in a believable way to (for them) weird behavior of other agents and the user.

- Serious games: For educational purposes, experience-based role-plays become possible, e.g. for increasing cultural awareness of users or for augmenting the standard language textbook with behavioral learning.

To provide conversational agents with a cultural background that will influence their verbal and non-verbal behavior, reliable data on such cultural heuristics is necessary. The literature often gives very general or anecdotal evidence (e.g. [9], [12], [25], [11], [2]) like, for instance, high contact cultures tend to interact with less space between the interlocutors [9] or people from Southern European cultures tend to use more gestures in interactions [25]. For realizing agent systems, more fine grained information is necessary, e.g. on the synchronisation of different modalities (see e.g. [20]). To this end, we propose a standardized corpus collection in different cultures for prototypical situations that will allow us to extract the relevant information. In this paper we present the CUBE-G corpus, which so far comprises interactions from the German and the Japanese culture.

RELATED WORK

Apart from data-driven algorithms, the use of annotated corpora has started to spread over from the social sciences to

computer science over the last ten years due to a number of different reasons. Basically, three types of corpus use can be distinguished nowadays:

1. Training data-driven algorithms: The use of language corpora has a long tradition in language processing for example to train speech recognizer or synthesizer. To this end, annotated corpora of spoken language are employed to train e.g. statistical classifiers like Naïve Bayes or Hidden Markov Models (see [8] for a concise overview).

2. Supporting information retrieval: Annotating large amounts of data with additional information establishes a large database that can be used to retrieve information based on the annotations. Examples for successful annotation schemes in this respect are the idv3 tags for mp3-files or the exif scheme for annotating image information. [14] for instance present a system to automatically annotate files that are created "on the move" with context information. This information is derived from devices that are available in a given situation like a temperature sensor and a GPS device which gather information about the latitude and longitude as well as the temperature in the environment at the moment a picture was taken.

3. Extracting empirical data on interaction behavior: Often data on human interaction is lacking information necessary for developing a model to control the behavior of a conversational agent (e.g. about the synchronization of different modalities). To keep the intuition of the reseacher at bay, it is indispensable to collect and annotate this data. Once create such a database can serve to extract rules or statistical information for behavior generation and analysis or it can serve as a benchmark against which the resulting system can be tested.

Especially the last point is interesting for enculturating interfaces and developing conversational agents with a cultural background. A number of large corpora of multimodal behavior already exist but all of them focus on the interaction without taking cultural difference of behavior into account. The AMI (Augmented Multipartiy Interaction) corpus[1] comprises around 100 hours of meeting recordings featuring verbal and nonverbal interactions between multiple interlocutors. A similar corpus was collected in the CHIL project[2] (Computer in the Human Interaction Loop). The Smartkom corpus[3] focuses on human computer interaction and was recorded in a Wizard of Oz setting to access users' interaction habit with a virtual character. The SAL corpus[4] (Sensitive Artificial Listener) is mainly concerned with investigating facial expressions of emotion. Because Ekman (e.g. [7]) has shown the existence of display rules for emotions that vary from culture to culture, it seems inevitable that the SAL corpus has to be augmented with recordings from different cultures.

Although embodied conversational agents are ideal candidates for integrating cultural aspects of interaction, there are few approaches that actually consider this challenge. [6] illustrate this problem by their survey of the Microsoft Agents web site which shows, that the appearance as well as the animations of the characters are all based on western cultural norms. [24] emphasizes this problem as a "McDonaldization" of agents, if culture-specific aspects are disregarded in the design and behavioral modeling of agents. Why this can create a problem has been shown by Nass and colleagues, who claim that the cultural background and behavioral consistency of an agent matter. In one of their studies, Korean subjects were confronted with either an American or a Korean agent. The subjects trusted the agent which corresponded to their own cultural identity more. From a technical point of view, the problem arises of how to ensure consistency between verbal and nonverbal communicative behaviors. An agent that just stares at the interaction partner and does not show any appropriate eye movements or gaze behavior will create an awkward atmosphere which may well lead to a failure of the interaction. To prevent such failures of communication and make agents believable and consistent in their behavior, the EMOTE model [1] seems to provide a promising starting point since it enables the generation of several variants for the same basic animation data depending on the settings of parameters, such as effort and shape. [18] define a specific markup language called GESTYLE which allows the user to vary an agent's style both for verbal and nonverbal modalities. This might be a possibility to integrate culture-specifc styles of behavior. In contrast to them, [16] employ a copy-synthesis approach to specify expressivity dimensions for an embodied conversational agent. The synthesis of culturally dependent behavior could be realized in their system by simply processing the corresponding videos from different cultures. [13] describe a language tutoring system that also takes cultural differences in gesture usage into account. The users are confronted with some prototypical settings and apart from speech input, have to select gestures for their avatars. Moreover they have to interpret the gestures by the tutor agents to solve their tasks. [5] describe a training scenario for different negotiation styles which is set in a different culture than the trainees'. Although this setting might be regarded as a prototypical case for rendering the system's behavior culture specific, especially regarding different types of negotiation, this aspect is not integrated in the system so far.

THE CUBE-G CORPUS
The rationale for creating the CUBE-G corpus was the lack of principled studies analyzing and comparing observational data from different cultures in a standardized way. Our starting point was Hofstede's dimensional model of culture that allows for unambiguously distinguishing given cultures on five dimensions.

1. Hierarchy: This dimension deals among other things, with superiors' decision-making styles and with the decision-making style that subordinates prefer in their boss. Hof-

---

2

Figure 1. German and Japanese actors during rehearsal for the three prototypical situations.

stede concludes that more coercive and referent power is used in high-H societies and more reward, legitimate, and expert power in low-H societies.

2. Identity: The degree to which individuals are integrated into a group is defined with this dimension. On the individualist side we find societies in which the ties between individuals are loose: everyone is expected to look after him/herself. On the collectivist side, we find societies in which people are integrated into strong, cohesive in-groups.

3. Gender: The gender dimension describes the distribution of roles between the genders. In feminine cultures the roles differ less than in masculine cultures, where competition is rather accepted and status symbols are of importance.

4. Uncertainty: The tolerance for uncertainty and ambiguity is defined in this dimension. It indicates to what extent a culture programs its members to feel either uncomfortable or comfortable in unstructured situations. Unstructured situations are novel, unknown, surprising, or different from usual.

5. Orientation: Values associated with long term orientation are thrift and perseverance wheras values associated with short term orientation are respect for tradition, fulfilling social obligations, and protecting one's face.

To ground the cultural influences on the behavior of an embodied conversational agent in hard empirical data, we devised a standardized observational study starting with two cultures that are located on different areas of the Hofstede dimensions, namely Germany and Japan. Three prototypical interaction scenarios were defined that are found in every culture to allow for comparing the verbal and nonverbal behavior (see Figure 1 for an impression).

1. Meeting someone for the first time: This is the standard first chapter of every language learning textbook and one of the most fundamental interactions in everyday communication.

2. Negotiating: Coming to an agreement with others can also be considered as a fundamental interaction esp. in intercultural communication. This scenario allows us to compare different negotiation styles and the accompanying verbal and nonverbal behavior.

3. Interacting with higher status individual: Cultures differ in how they interpret the unequal distribution of power and status among the members of the culture, resulting in quite different behaviors towards interaction partners with a higher status.

These scenarios have been chosen due to two reasons. First of all, we claim that they represent situations every expatriate and even every tourist might easily encounter. Moreover, we expect different verbal and/or nonverbal behavior patterns in the German and the Japanese culture due to their different locations on Hofstede's dimensions. Following Hofstede, we expect differences in persuasion strategies based on the hierarchy dimension (see Introduction). Additionally, we expect differences related to the identity dimensions, i.e. we expect Japanese participants to use more in-group markes than German participants.

Design of the Study
Dyadic interactions between human subjects were recorded in the three scenarios mentioned above. Table 1 gives an

3

First time meeting		Negotiation		Social status	
Actor	Subjects	Actor	Subjects	Actor	Subjects
M_{A1}	M_{S1}-M_{S5}	M_{A1}	M_{S1}-M_{S5}	M_{A2}	M_{S1}-M_{S5}
	F_{S1}-F_{S5}		F_{S1}-F_{S5}		F_{S1}-F_{S5}
F_{A1}	M_{S6}-M_{S10}	F_{A1}	M_{S6}-M_{S10}	F_{A2}	M_{S6}-M_{S10}
	F_{S6}-F_{S10}		F_{S6}-F_{S10}		F_{S6}-F_{S10}

Table 1. Design corpus study

overview of the design. One of the interaction partners in each scenario was an actor following a script for the specific situation. The rationale for using actors as interaction partners was that we would be able to elicit sufficient interactions from the subjects and to control the conditions for each participant more tightly. To control for gender effects, a male and a female actor were employed in each scenario interacting with the same number of male and female subjects. Thus, apart from the two male (MA1, MA2) and two female actors (FA1, FA2), at least ten male (MS1-MS10) and ten female subjects (FS1-FS10) were needed for this corpus study. The same design was used in Germany as well as in Japan.

Actors were supplied with background stories according to their role. The backstory of participant actors took most of their personal history into account (where they lived, their hobbies, etc). Additionally they got information about the subject they were supposed to be studying. For the first meeting scenario, actors were told to mainly react to the participant and only if the conversation got stuck to take the initiative. For the negotiation task, actors were supplied with a pro-/con-list for the items. They were told to initally agree only on one item with the participant. Thus, every pair had to negotiate about the remaining four items. High status actors were representatives from a large consulting company that conducted the study, and had to debrief the participants. To this end, they had the "official" list and were also supplied with a pro-/con-list for the items.

The actual number of participants differed between Germany and Japan. 21 subjects (11 male, 10 female) participated in the German data collection, 26 subjects (13 male, 13 female) in the Japanese collection. For each subject, around 25 minutes of video material was collected, 5 minutes for the first meeting, 10-15 minutes for the negotiation, and 5 minutes for the status difference. Participants were told that they take part in a study by a well-known consulting company for the automobile industry, which would take place at the same time in different countries. To attract their interest in the study, a monetary reward was granted depending on the outcome of the negotiation task.

To ensure equal conditions on both sides, a recording booth was designed that features two video cameras, a web cam and a micorphone. The booth itself was around 3x3m. Each of the video cameras was recording one of the participants (actor and subject), and was barely visible from the inside. The microphone was installed between participants at the side of the booth to ensure a good recording quality on the one hand, but also no interruption with the spatial behavior

of the subjects on the other hand. The webcam was installed on the floor at the side of the booth hidden from view. It recorded the proxemics behavior of the participants, i.e. how close they move to each other during the interaction.

During the experiment, the participant first had to fill out a NEO-FFI personality questionnaire [17]. Afterwards he was led into the recording room, where the actor who posed as student was already waiting. Both were greeted by the high status actor who posed as representative for the consulting company. They were told that they had five minutes to get to know each other which would make the negotiation faster easier. During this time the equipment would be checked. Of course, this first meeting was recorded. Afterwards the negotiation task was presented, which is a variation of the standard "Lost-at-sea" scenario. Participants were told that they are on a boat trip in the South Pacific when their boat is sinking. They can grap three items from a 15 item list that will help them survive and get rescued. During their negotiation they had to come up with a single three item list where the items are ranked in order of importance. This scenario was chosen because it offers the possibility of using quite different negotiation and persuasion strategies. For instance participants could argue in an integrative way taking all arguments into account or they could try to get their own items on the list regardless of arguments. After the negotiation, the student actor was shown out of the room, the high status actor started debriefing the participant who had to explain the rationales for the selected items and their ranking. Depending if their items and the ranking were consistent with the "official" list by the U.S. coast guard they would receive a monetary reward of up to 30 Euros (3000 Yen). To elicit behavior for pre-defined situations, the actor followed a script. First, a positive atmosphere was created by agreeing with the participant on the first item. The second item, was in the top three but not on the second place. The third item was completely wrong. The actor does not offer explanations himself, but asked the participant for the reasons of picking the items. Thus, we were able to see if participants would question the high status interlocutor by letting him explain the reasons for the ranking on the "official" list or even start arguing about this "official" ranking.

COMPARATIVE CORPUS ANALYSIS

The analysis of the CUBE-G corpus is concentrating on nonverbal behavior at the moment. The behavior under investigation is comprised of postures, gestures, gestural expressivity, gaze, volume, and proxemics. Of course the corpus allows for unlimited further comparisons of behavior in the

4

		GM	JP	t
Frequency	Head	22	15.6	n.s.
	Leg	9.5	16.56	+
	Arm	40.38	22.8	+
	All	71.88	58.56	n.s
Duration	Head	2.57	2.54	n.s.
	Leg	19.93	24.64	n.s.
	Arm	7.79	14.08	+

$+ : p < 0.1, * : p < 0.05, ** : p < 0.01$

Table 2. Posture shift frequency and duration

two cultures.[5] In this article we concentrate on our first results on differences in posture and expressivity.

Differences in Posture

We employ Bull's posture coding scheme [4] to categorize posture shifts observed in our corpus. For the current analysis, we annotated head, arm, and leg postures for 8 German and 9 Japanese first time meeting conversations. The following types of behaviors in Bull's categories were observed in these interaction data.

Head postures
RaUHd: Raises up head
DsHd: Drops head
LHdF: Leans head forward
LHdP: Leans head to person
THdAP: Turn head away from person
LHdAP: Leans head away from person

Leg postures
MLP: Moves leg to person
MLAP: Moves leg away from person
LSF: Leans sideways on foot
Xls: Crosses legs

Arm postures
(a) Hand-to-head postures
PHHd: Puts hand to head
PHNk: Puts hand to neck
PHFe: Puts hand to face
(b) One-handed postures
PHSr: Puts hand to shoulder
PHUAm: Puts hand to upper arm
PHEw: Puts hand to elbow
PHLAm: Puts hand to lower arm
PHWr: Puts hand to wrist
(c) Two-handed postures
FAs: Folds arms
JHs: Joins hands
(d) Hand-to-trunk postures
PHCt: Puts hand to chest
PHAn: Puts hand to abdomen
PHB: Puts hand to back
(e) Hand-to-clothes postures
HP: Holds pullover
PHIPt: Puts hand into pocket
PHTr: Puts hands to trousers

---

[5]Proposals for joint efforts are always welcome. Please have a look at http://mm-werkstatt.informatik.uni-augsburg.de/projects/cube-g/ for contact details.

Head posture analysis
The results of comparison for the frequency and the duration of head, arm, and leg postures were shown in Table 2. The average number of head posture shifts in German data was 22 and that in Japanese data is 15.6. The average duration of each posture (how long the subjects were keeping the same posture) in German data was 2.57 and that in Japanese data was 2.54. As the results of conducting t-tests, these differences were not statistically significant (t(16)= 1.095; ns, and t(16)= 0.0562; ns, respectively).

However, the distribution of the categories was different between the two cultures. As shown in Fig. 2, Japanese people generally did less head posture shifts than German people, except for THdAP. The difference of distribution was statistically significant ($\chi^2(5) = 20.308, p < 0.05$).

Leg posture analysis
The average number of leg posture shifts in German data was 9.5 and that in Japanese data was 16.56. A weak trend was found in a t-test (t(15)= 1.764, $p < 0.1$). The average duration of each posture in German data was 19.93 and that in Japanese data was 24.64 (Table 2), but the difference was not statistically significant (t(15)= 0.409, ns). As shown in Fig. 3, in the distribution analysis, we found that the difference in category distribution was statistically significant ($\chi^2(3) = 9.205, p < 0.05$). While LSF were the most frequent in both German and Japanese data, Japanese people also frequently did MLP.

Arm posture analysis
As shown in Table 2, the average number of arm posture shifts in German data was 40.38 and that in Japanese data was 22.8. A weak trend was found in a t-test (t(16)= 1.931, $p < 0.1$). The average duration of each posture in German data was 7.79 and that in Japanese data was 14.08. We also found a trend for this comparison (t(16)= 2.061, $p < 0.1$). In distribution analyses, we analyzed the distribution for each sub-category: (a) hand-to-arm (one-handed), (b) hand-to-arm (two-handed), (c) hand-to-head, (d) hand-to-trunk, and (e) hand-to-cloth (Fig. 4 (a)-(e)). The differences in category distributions were statistically significant in hand-to-arm (one-handed), hand-to-arm (two-handed), hand-to-head, and hand-to-cloth postures ($\chi^2(4) = 70.482; p < 0.01$, Fisher's Exact Test $p < 0.01$, $\chi^2(2) = 7.208; p < 0.01$, and $\chi^2(2) = 91.447; p < 0.01$, respectively), and a trend was found in hand-to-trunk ($\chi^2(2) = 5.708, p < 0.1$). Hand-to-head postures more frequently occurred in Japanese data than German data, especially PHFe was the most frequent in Japanese data. Hand-to-arm (one-handed) postures were very different depending on the culture. The most frequent category in German data was PHEw, and that in Japanese data was PHWr. Interestingly, German people rarely did PHWr, and Japanese people rarely did PHEw. As for hand-to-arm (two-handed) postures, the most frequent category in German data was FAs and that in Japanese data was JHs. Hand-to-cloth postures were rarely observed in Japanese data, but, especially PHIPt, they were very frequent in German data.
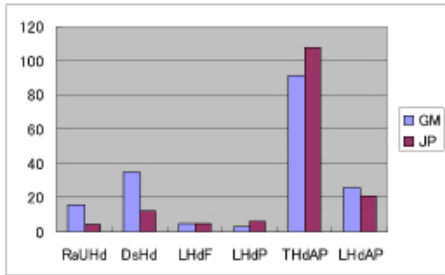
5

Figure 2. Head posture distribution
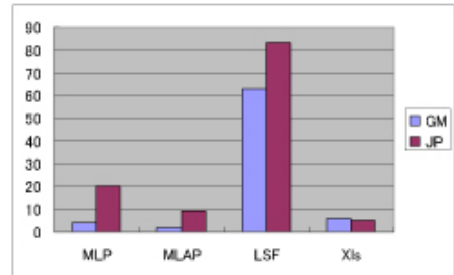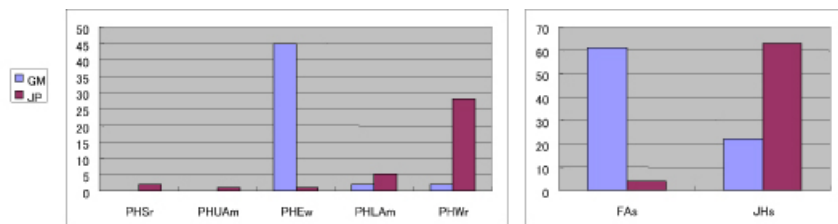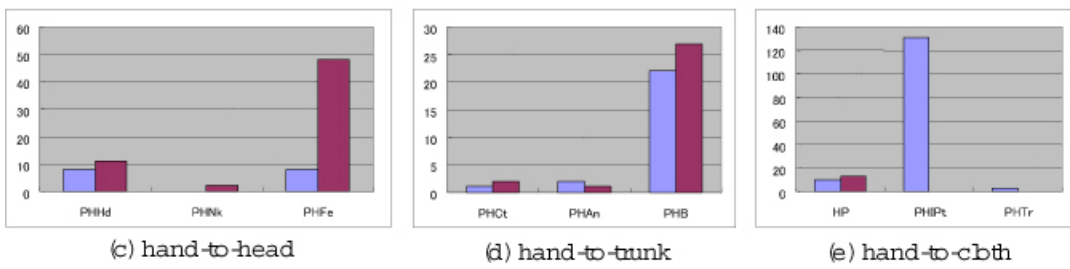

Figure 3. Leg posture distribution

(a) one-handed

(b) two-handed

(c) hand-to-head

(d) hand-to-trunk

(e) hand-to-cbth

Figure 4. Arm posture distribution

6

Discussion of posture analysis
Generally, head postures were not very different depending on the culture, but Japanese people more frequently looked away from the partner than German people. Cultural difference was clearer in arm postures. German people more frequently changed arm postures than Japanese people, and Japanese people kept the same posture longer than German. Posture shapes were also very different. German people mainly used their arms, such as folding their arms (FAs) and put their hands on the elbows (PHEw). On the contrary, Japanese people mainly used their hands, such as joining the hands (JHs), putting their hands on the wrists (PHWr). Moreover, Japanese people frequently touched their heads by their hands, and German people put their hands in the pockets. Although Japanese people did not move their upper bodies as frequently as German people, they used more leg postures.

In addition to these results above, we also found that the total number of posture shifts per conversation is not different depending on the culture (71.88 in German data and 58.56 in Japanese data, and the difference is not statistically significant (t(15)= 1.154, ns)). All these results suggest that the frequency of posture shifts is not different depending on the culture, but the types of postures frequently used are different. Thus, the body parts used as well as shapes of the postures express the characteristics of each culture.

Differences in Gestural Expressivity

The coding scheme and the analysis of gestural expressivity follows [19]. From the available material so far, eight interactions for the first meeting scenarios have been annotated for each culture. Gestural expressivity was coded for the five parameters repetition, fluidity, power, speed, and spatial extent. Each parameter was coded using a seven-point scale where 1 denotes small values and 7 large values for the parameter (except for repetition where it denotes the number of repetitions of a given gesture). The distinction between power and speed is taken over from [3]. In order to gain insights in the supposed differences in the use of gestures, we compared expressivity parameters of the German and the Japaneses samples. Moreover, we looked into gender specific differences. First of all, it has to be said that there is a noticeable difference in the number of gestures that were used in the German and the Japanese samples. German participants used three times more gestures than Japanese participants (177 vs. 53). Table 3 gives the results for this analysis (ANOVA). For the overall comparison between the German and the Japanese sample, no significant difference can be seen for the parameter power and only a weak trend for speed ($p < 0.1$). For the other parameters the difference is highly significant ($p < 0.01$). When we compared only the male samples, the results were slightly different. There was no significant difference for spped but a weak trend for power ($p < 0.1$). Comparison of the female samples from the two cultures yielded a different picture although no significant differences are present between genders in a given culture. Expressivity of females does not vary as much as expressivity of male participants. The only significant differences can be found for the fluidity parameter ($p < 0.01$)

and the speed parameter ($p < 0.05$), and there is a weak trend for repetition ($p < 0.1$).

Discussion of gestural expressivity
These preliminary results show a tendency concerning the differences in how gestures are expressed in the two cultures. But as we have only looked into part of the available corpus it remains to be shown that the results are also stable for a larger sample. The most interesting result concerns the gender specific differences in gesture use between the two cultures. Whereas the differences for the male participants are strong and highly significant, females from the two cultures do not statistically differ much in their gesture use. Table 3 shows that the means for the females of a given culture are located towards the means for the males of this culture for each parameter.

Interestingly, spatial extent is rated higher for the Japanese sample. Now this could be an interesting side effect of culture specific interpretations. So far, the Japanese samples have been labeled by Japanese and the German samples have been labeled by Germans. Although the coders were given a coding manual that specified how to label the spatial extent based on elbow contraction, it might still be the case that culture specific interpretation might have taken place. Thus, it might be interesting to let the Japanese coders label the German data and vice versa as a next step.

CONCLUSION

In this paper we presented a corpus of multimodal behavior that was collected under standardized conditions for three prototypical conditions in two different cultures, Germany and Japan. It was argued, that such a principled approach is needed to endow conversational agents with culture-specific verbal and nonverbal behavior which will further the successful use of such agent systems in the area of information presentation, persuasion, and edutainment. A preliminary analysis of the corpus was presented that focused on specific nonverbal aspects of communication, body posture and gestural expressivity. For both aspects of behavior, differences between the cultures were found on different levels of granularity.

Body posture as well as gestural expressivity are not only determined by one's cultural background. Indeed, the cultural background only gives general behavioral heuristics which might e.g. result in preferring higher spatial extent. But such behaviors are also dependent on personality or personal style. This was not taken into account in the preliminary analysis presented here. To test for influences of personality on observed behavior, every participant had to do a NEO-FFI personality test [17]. The results from these tests will allow us to analyse correlations between personality traits of our participants and behavior patterns. Moreover, the corpus collection was designed to register differences in the interaction between members of the same sex and between members of different sexes. In the above mentioned results, these different gender constellations were not taken into account due to the fact, that annotation started with random samples and we only tried to have the same number of female and male participants annotated.

7

	All			Male			Female		
Parameter	JP	G	F	JP	G	F	JP	G	F
Repetition	2.70	1.38	50.247**	2.90	1.34	34.173**	2.00	1.40	3.537$^+$
Fluidity	2.87	4.40	68.591**	2.83	4.2	68.544**	3.00	4.52	12.379**
Power	3.39	3.21	0.736	3.55	3.08	3.399$^+$	2.83	3.28	1.092
Speed	3.39	3.81	2.929$^+$	3.60	3.52	0.780	2.67	3.98	6.499*
Sp.Ext.	4.02	3.01	18.703**	4.14	2.89	21.609**	3.58	3.08	1.066
Duration	5.13	2.39	15.461**	5.83	1.98	20.192**	2.67	2.63	0.001

$^+ : p < 0.1$, $^* : p < 0.05$, $^{**} : p < 0.01$

Table 3. Results of expressivity analysis

The information from this corpus study has been integrated in the behavior generation process of an embodied conversational agent system. [22] gives an account of the system architecture where culture penetrates all processing steps from the planning of the utterance to selecting and modifying appropriate animations. To this end, the system makes use of Bayesian network that models the causal relations between cultural dimensions and behavior ([21]). The probability distributions of this network can now be grounded in the empirical data about gestural expressivity. In a pilot study it was shown that users prefer agents that adhere to the behavior related to their own cultural background.

REFERENCES
1. J. M. Allbeck and N. I. Badler. Creating Embodied Agents With Cultural Context. In S. Payr and R. Trappl, editors, *Agent Culture: Human-Agent Interaction in a Multicultural World*, pages 107–126. Lawrence Erlbaum Associates, London, 2004.

2. M. Argyle. *Bodily Communication*. Methuen & Co. Ltd., London, 1975.

3. E. Bevacqua, A. Raouzaiou, C. Peters, G. Caridakis, K. Karpouzis, C. Pelachaud, and M. Mancini. Multimodal sensing, interpretation and copying of movements by a virtual agent. In *PIT*, pages 164–174, 2006.

4. P. E. Bull. *Posture and Gesture*. Pergamon Press, 1987.

5. M. Core, D. Traum, H. C. Lane, W. Swartout, J. Gratch, M. V. Lent, and S. Marsella. Teaching negotiation skills through practice and reflection with virtual humans. *SIMULATION*, 82(11):685–701, 2006.

6. F. de Rosis, C. Pelachaud, and I. Poggi. Transcultural Believability in Embodied Agents: A Matter of Consistent Adaptation. In S. Payr and R. Trappl, editors, *Agent Culture: Human-Agent Interaction in a Multicultural World*, pages 75–106. Lawrence Erlbaum Associates, London, 2004.

7. P. Ekman. *Telling Lies — Clues to Deceit in the Marketplace, Politics, and Marriage*. Norton and Co. Ltd., New York, 3rd edition, 1992.

8. M. E. Foster. Issues for corpus-based multimodal generation. In *Proceedings of the Workshop on Multimodal Output Generation (MOG 2007)*, pages 51–58, 2007.

9. E. T. Hall. *The Hidden Dimension*. Doubleday, 1966.

10. G. Hofstede. *Cultures Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Sage Publications, Thousand Oaks, London, 2001.

11. G. J. Hofstede, P. B. Pedersen, and G. Hofstede. *Exploring Culture: Exercises, Stories, and Synthetic Cultures*. Intercultural Press, Yarmouth, 2002.

12. K. Isbister. Building Bridges Through the Unspoken: Embodied Agents to Facilitate Intercultural Communication. In S. Payr and R. Trappl, editors, *Agent Culture: Human-Agent Interaction in a Multicultural World*, pages 233–244. Lawrence Erlbaum Associates, London, 2004.

13. L. Johnson, R. Mayer, E. André, and M. Rehm. Cross-Cultural Evaluation of Politeness in Tactics for Pedagogical Agents. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED)*, 2005.

14. A. Karypidis and S. Lalis. Automated context aggregation and file annotation for PAN-based computing. *Personal and Ubiquitous Computing*, 11(1):33–44, 2007.

15. A. Marcus. Cultural Dimensions and Global Web User-Interface Design: What? So What? Now What? In *HFWEB2000 Conference Proceedings*, 2000.

16. J.-C. Martin, S. Abrilian, L. Devillers, M. Lamolle, M. Mancini, and C. Pelachaud. Levels of Representation in the Annotation of Emotion for the Specification of Expressivity in ECAs. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist, editors, *Intelligent Virtual Agents*

8

(IVA), pages 405–417, Berlin, Heidelberg, 2005. Springer.

17. R. R. McCrae and O. P. John. An introduction to the five factor model and its applications. *Journal of Personality*, (60):175–215, 1992.

18. H. Noot and Z. Ruttkay. Variations in Gesturing and Speech by GESTYLE. *International Journal of Human-Computer Studies, Special Issue on Subtle Expressivity for Characters and Robots*, 2005.

19. C. Pelachaud. Multimodal expressive embodied conversational agents. In *Proceedings of ACM Multimedia*, pages 683–689, 2005.

20. M. Rehm and E. André. From annotated multimodal corpora to simulated human-like behavior. In I. Wachsmuth and G. Knoblich, editors, *Modeling Communication with Robots and Virtual Humans*. Springer, in press.

21. M. Rehm, N. Bee, B. Endrass, M. Wissner, and E. André. Too close for comfort? Adapting to the user's cultural background. In *Workshop on Human-Centered Multimedia, ACM Multimedia*, 2007.

22. M. Rehm, Y. Nakano, E. André, and T. Nishida. Culture-specific first meeting encounters between virtual agents. In H. Prendinger et al., editors, *Intelligent Virtual Agents*. Springer, 2008.

23. S. M. Schmidt and R. S. Yeh. The structure of leader influence: A cross-national comparison. *Journal of Cross-Cultural Psychology*, 23:251–264, 1992.

24. P. Sengers. The Agents of McDonaldization. In S. Payr and R. Trappl, editors, *Agent Culture: Human-Agent Interaction in a Multicultural World*, pages 3–20. Lawrence Erlbaum Associates, London, 2004.

25. S. Ting-Toomey. *Communicating Across Cultures*. The Guilford Press, New York, 1999.

9