Vorlesung Mensch-Maschine-Interaktion

Albrecht Schmidt
Embedded Interaction Research Group
LFE Medieninformatik
Ludwig-Maximilians-Universität München
http://www.hcilab.org/albrecht/

Albrecht Schmidt
Embedded Interaction Research Group
University of Munich, Germany

MMI 2005/2006

Chapter 7
Evaluation
(selected topics)

Further Issues

- **Consent form – get written consent from participants**
  - Templates available on the Internet
  - May be checked with the legal department / review board

- **Let participants know what they are doing**
  - What is the participant expected to do
  - Procedure
  - How long will it take, breaks
  - What is the study for in general – but do NOT tell about the specific purpose or your hypotheses

- **Make sure they know**
  - Quality of a UI / software is tested
  - They are NOT tested

- **Ethical Issues**

Albrecht Schmidt
Embedded Interaction Research Group
University of Munich, Germany

MMI 2005/2006

Participants Consent (Example)

Participants Consent Form

Study _____ **Institution** _____

Name: _____ Date of Birth: _____

Email: _____

Phone:_____

I have been informed on the procedure and purpose of the study and my questions have been answer to my satisfaction.

I have volunteered to take part in this study and agree that during the study information is recorded (audio and video as well as my interaction with the system). This information may only be used for research and teaching purpose. I understand that my participation in this study is confidential. All personal information and individual results will not be released to third parties without my written consent.

I understand that I can withdraw from participation in the study at any time.

Date: _____ Signature:_____

Chapter 7
Evaluation
(selected topics)

- 7.1 User studies
- **7.2 Heuristic Evaluation**
- 7.3 Minimal Usability Lab
- 7.4 Cognitive Walkthrough
- 7.5 Questionnaires and Interviews

Why evaluate?
Goals of user interface evaluation

- Ensure functionality (effectiveness)
  - Assess (proof) that a certain task can be performed
- Ensure performance (efficiency)
  - Assess (proof) that a certain task can be performed given specific limitations (e.g. time, resources)
- Customer / User acceptance
  - What is the effect on the user?
  - Are the expectations met?
- Identify problems
  - For specific tasks
  - For specific users
- Improve development life-cycle
- Secure the investment (don't develop a product that can only be used by fraction of the target group – or not at all!)

There is not a single way ...

- **Different approaches**
  - Inspections
  - Model extraction
  - Controlled studies
  - Experiments
  - Observations
  - Field trails
  - Usage context

- **Different results**
  - Qualitative assessment
  - Quantitative assessment

Usability Methods are often not used!

- **Why**
  - Developers are not aware of it
  - The expertise to do evaluation is not available
  - People don't know about the range of methods available
  - Certain methods are to expensive for a project (or people think they are to expensive)
  - Developers see no need because the product "works"
  - Teams think their informal methods are good enough

- **starting points**
  - Discount Usability Engineering
    http://www.useit.com/papers/guerrilla_hci.html
  - Heuristic Evaluation
    http://www.useit.com/papers/heuristic/

Inspections & Expert Review

- Throughout the development process
- Performed by developers and experts
- External or internal experts
- Tool for finding problems
- May take between an hour and a week
- Structured approach is advisable
  - reviewers should be able to communicate all their issues (without hurting the team)
  - reviews must not be offensive for developers / designers
  - the main purpose is finding problems
  - solutions may be suggested but decisions are up to the team

Inspection and Expert Review Methods

- **Guideline review**
  - Check that the UI is according to a given set of guidelines

- **Consistency inspection**
  - Check that the UI is consistent (in itself, within a set of related applications, with the OS)
  - Birds's eye view can help (e.g. printout of a web site and put it up on the wall)
  - Consistency can be enforced by design (e.g. css on the web)

- **Walkthrough**
  - Performing specific tasks (as the user would do them)

- **Heuristic evaluation**
  - Check that the UI violates a set (usually less than 10 point) rules

Informal Evaluation

- Expert reviews and inspections are often done informally
  - UIs and interaction is discussed with colleagues
  - People are asked to comment, report problems, and suggest additions
  - Experts (often within the team) assess the UI for conformance with guidelines and consistency
- Results of informal reviews and inspections are often directly used to change the product
- … still state of the art in many companies!
- Informal evaluation is important but in most cases not enough

- Making evaluation more explicit and documenting the findings can increase the quality significantly
- Expert reviews and inspections are a starting point for change

Discount Usability Engineering

- Low cost approach
- Small number of subjects
- Approximate
  - Get indications and hints
  - Find major problems
  - Discover many issues (minor problems)
- Qualitative approach
  - observe user interactions
  - user explanations and opinions
  - anecdotes, transcripts, problem areas, …
- Quantitative approach
  - count, log, measure something of interest in user actions
 - speed, error rate, counts of activities

Heuristic Evaluation

http://www.useit.com/papers/heuristic/

- Heuristic evaluation is a usability inspection method
- systematic inspection of a user interface design for usability
- goal of heuristic evaluation
  - to find the usability problems in the design
- As part of an iterative design process.

- Basic Idea:
  Small set of evaluators examine the interface and judge its compliance with recognized usability principles (the "heuristics").

Heuristic Evaluation

http://www.useit.com/papers/heuristic/

- How many evaluators?
- Example: total cost estimate with 11 evaluators at about 105 hours, see
  http://www.useit.com/papers/guerrilla_hci.html
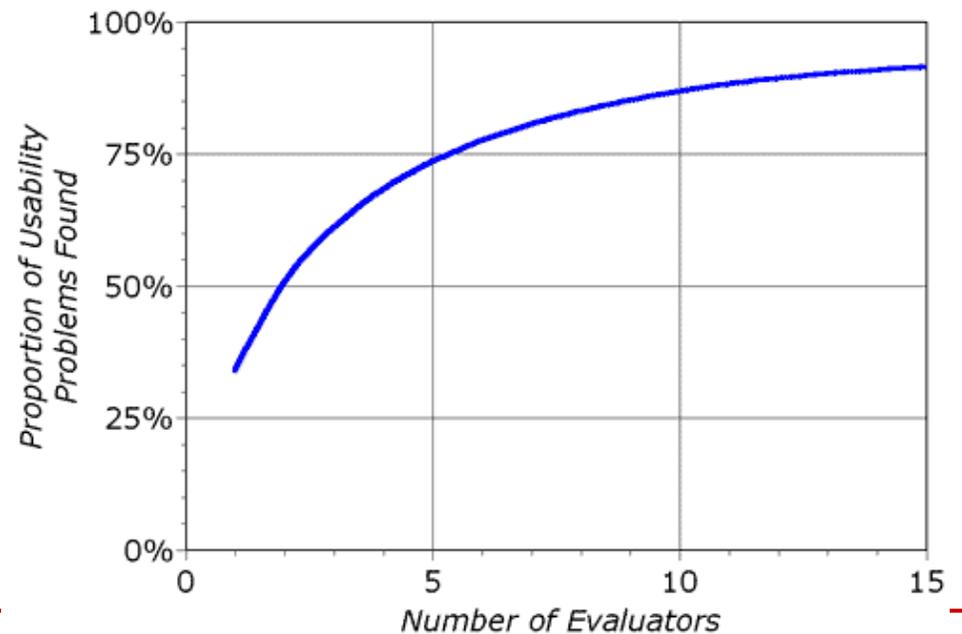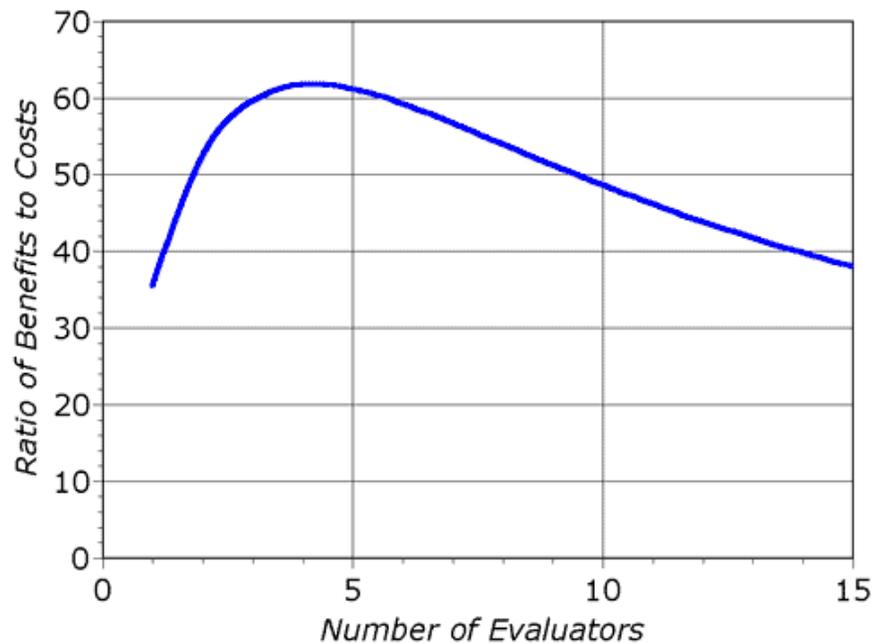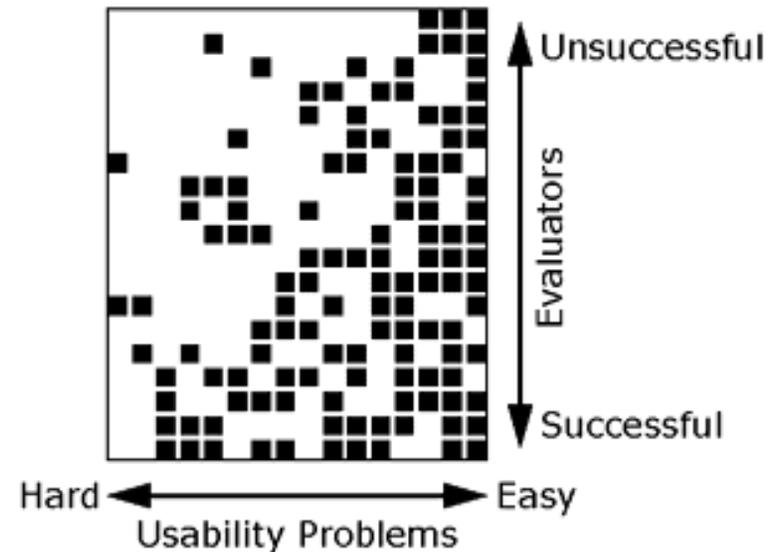
Heuristic Evaluation - Heuristics

- Heuristics suggested by Nielsen
  - Visibility of system status
  - Match between system and the real world
  - User control and freedom
  - Consistency and standards
  - Error prevention
  - Recognition rather than recall
  - Flexibility and efficiency of use
  - Aesthetic and minimalist design
  - Help users recognize, diagnose, and recover from errors
  - Help and documentation

- Depending of the product and goals a different set may be appropriate

Heuristic Evaluation - Steps

- Preparation
  - Assessing appropriate ways to use heuristic evaluation
  - Define Heuristics
  - Having outside evaluation expert learn about the domain and scenario
  - Finding and scheduling evaluators
  - Preparing the briefing
  - Preparing scenario for the evaluators
  - Briefing (system expert, evaluation expert, evaluators)
  - Preparing the prototype (software/hardware platform) for the evaluation
- Evaluation
  - Evaluation of the system by all evaluators
  - Observing the evaluation sessions
- Analysis
  - Debriefing (evaluators, developers, evaluation expert)
  - compiling list of usability problems (using notes from evaluation sessions)
  - Writing problem descriptions for use in severity-rating questionnaire
  - Severity rating

Heuristic Evaluation – Severity Rating

- Severity ratings are used to prioritize problems
- Decision whether to release a system or to do further iterations
- The severity of a usability problem is a combination of three factors:
  - The frequency with which the problem occurs: Is it common or rare?
  - The impact of the problem if it occurs: Will it be easy or difficult for the users to overcome?
  - The persistence of the problem: Is it a one-time problem that users can overcome once they know about it or will users repeatedly be bothered by the problem

- 0 to 4 rating scale to rate the severity of usability problems:
  - 0 = I don't agree that this is a usability problem at all
  - 1 = Cosmetic problem only: need not be fixed unless extra time is available on project
  - 2 = Minor usability problem: fixing this should be given low priority
  - 3 = Major usability problem: important to fix, so should be given high priority
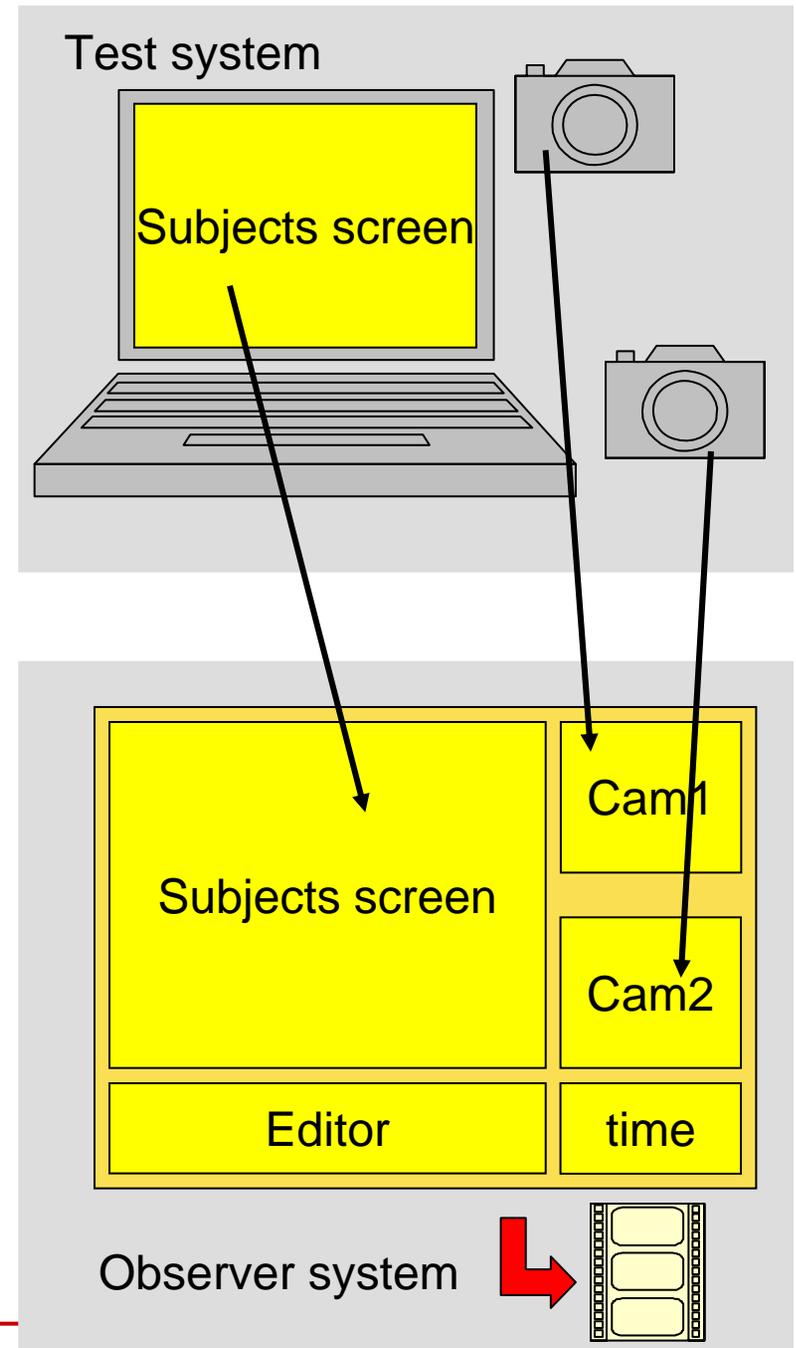 - 4 = Usability catastrophe: imperative to fix this before product can be released

Chapter 7
Evaluation
(selected topics)

- **7.1 User studies**
- **7.2 Heuristic Evaluation**
- **7.3 Minimal Usability Lab**
- **7.4 Cognitive Walkthrough**
- **7.5 Questionnaires and Interviews**

Video protocol

- Integrate multiple views
  - Capture screen with pointer
  - View of the person interacting with the system
  - View of the environment

- Poor man's usability lab
  - Computer for the test user,
    - run application to test
    - export the screen (e.g. VNC)
  - Computer for the observer
    - See the screen from the subject
    - Attach 2 web cams and display them on the screen
    - Have an editor for observer notes
    - Capture this screen (e.g. camtasia)

- Discuss with the user afterwards
  - Why did you do this?
  - What did you try here?
  - ….

Test system

Subjects screen

Subjects screen

Cam1

Cam2

Editor

time

Observer system

Albrecht Schmidt
Embedded Interaction Research Group
University of Munich, Germany

MMI 2005/2006

Screen video

Chapter 7
Evaluation
(selected topics)

Albrecht Schmidt
Embedded Interaction Research Group
University of Munich, Germany

MMI 2005/2006

Cognitive Walkthrough
Dix et al. Chapter 11

- For interfaces that can be learned by exploration
- Experts step through a task to question the design
  - focusing on the users' knowledge and goals
  - asking whether the users will experience difficulties at each step

- Requirements
  - A description of the system prototype i.e where will it be located, exact wordings of menus or a prototype.
  - A description of the task the user will be expected to do - the most common
  - A list of the correct actions that are required to complete the task
  - A description of who the users will be, their experience and prior knowledge

Cognitive Walkthrough - Questions
Dix et al. Chapter 11

- Evaluator works through the action list and at each step they ask:

  1. Will users be trying to produce whatever effect the action has?

  2. Will users see the control (button, menu, switch, etc.) for the desired action?

  3. Once users find the control, will they recognize that it produces the effect they want?

  4. After the action is taken, are the users given adequate feedback, so they can go on to the next action with confidence?

Organizing a Cognitive Walkthrough
Dix et al. Chapter 11

- **Requires good and precise documentation**
  - task description
  - details on action steps
  - user information

- **For each action step the evaluator comments of the four questions**

- **If the answer to any question is no, this indicates a usability problem → create a separate report**

- **For each problem found the evaluator should give a severity rating (helps to set priorities)**

Chapter 7
Evaluation
(selected topics)

Questionnaires and Interviews

- Lot of information available in psychology, communication studies, market research
- …here is just a quick overview

- Process to get the "right" questions
  - Brainstorm (within the project/design team) on issues that are relevant and should be put to the users
  - Select the set of relevant questions (make the size appropriate, don't ask questions you are not interested in)
  - Create a first version of the questionnaires or interviews
  - Run a few pilot interviews/questionnaires
  - Discuss the answers/results given – did participants understand what you wanted to ask them?
  - Potentially redesign the questionnaires or interviews
  - Run the interviews/questionnaires

Interviews

- Find out about users viewpoint
- Level of detail is not predetermined
- Allows more explanation and going into detail
- Open ended questions
- Good for exploration
- Often very dependent on the interviewer

- How to interview
  - Prepare a set of questions (core set for some consistency)
  - Ask question neutral and do not imply answers
    - "what is your opinion on the audio feedback" vs. "did you think the use of the audio feedback was really helpful"
- Group interviews
  - More discussion style
  - Finding a consensus
  - Often only the opinion of a few people in the group

Interviews

- **Recognize the users response**

- **Problem**
  - Time consuming
  - Interviewer can "steer" the outcome

- **Examples**
  - Retrospective interview after a test session
    - Show video recording and ask questions
    - Ask questions to clarify situations
  - Critical incident interviews
    - Ask about critical situation related to the software product
    - Rare events that may still be important

Questionnaires/Surveys

- To reach larger groups
- Initial effort may be large (creating the questionnaire and the analysis function)
- Creating them online (or at least machine readable) saves time
- Little effort per participant after the questionnaire is created
- Good for statistical analysis of results

… however if the questions are not good or the participants responding are the wrong ones the results may be poor

Questionnaires/Surveys

- How to create a questionnaire
  - Find out what the information is that you are you interested in
  - What should be analyzed and how should it be analyzed
  - What will the results be used for (e.g. redesign, new product, new features)

- Who is the audience
  - Specify the audience for questionnaire
  - How will representative participants be found

- What technology / approach will be used
  - Online / Webpage
  - Software
  - Paper

Style of Questions

- General
  - Explorative
  - Establish background
- Open ended questions
  - Set of answers are not pre-determined
  - Ask for opinion or subjective general comments
  - E.g. "what would you like to have different change on this web page"
  - Very hard to analyze automatically
- Closed questions
  - Types
    - Scalar
    - Ranked
    - Alternatives
    - Multiple choice
  - Response is restricted to alternatives
  - can be easily analyzed

- sometimes combined
  - "how did you hear about us? – TV, Radio, Google, other _____ "

Closed Questions
be specific

- Minimize interpretation for responses!
  - alternative answers should be very specific

    how often do you use computers at work:
    - ⊗ frequently
    - O sometimes
    - O rarely

    vs

    how often do you use computers at work on a typical work day
    - O more than 6 hours a day
    - O between 1 and 6 hours a day
    - ⊗ less than 1 hr a day

- For closed questions you must cover all sensible answers
- Watch the language (clear, avoid jargon)

Question Formats Scalar

- **Odd number allow neutral value**

  I found the audio feedback annoying
  Disagree 1   2   ③   4   5   Agree

- **even number forces a choice**

  I found the audio feedback annoying
  Disagree 1   2   3   ④ 5   6 Agree

- **Likert scale. 1-to-5 (or 1-7, 1-9)**

  1. strongly disagree
  2. disagree
  3. undecided
  4. agree
  5. strongly agree

Question Formats
Ranked

- As participants to rank options
- Example

  What method did you use most often to print the document?
  Please rank 1=most often, 2=middle, 3=least often

  Keyboard [3]   Toolbar [1]   Menu [2]

- Forces a choice on the participants

Albrecht Schmidt
Embedded Interaction Research Group
University of Munich, Germany

MMI 2005/2006

Question Formats
Alternatives & Multi Choice

- **Alternatives**
  - Give different options – but only one can be selected
  - Example
    what is your preferred way for electronic communication?
    [X] Email
    [ ] Fax
    [ ] SMS
    [ ] Video conferencing

- **Multiple choice**
  - Give different options – allow to select multiple of them
  - Example
    what forms of electronic communication have you use in the last 6 weeks?
    [X] Email
    [X] Fax
    [X] SMS
    [X] Video conferencing

Examples of methods used in different phases

- Analysis
  - Goal & user analysis
  - Task analysis
  - Contextual enquiry and observations

- Early design phase
  - Sketches and paper prototypes
  - Cognitive walkthroughs
  - Heuristic evaluation

- Late design phase
  - Functional prototypes
  - User studies and experiments

- Implementation
  - User studies
  - Functional tests
  - Acceptance tests
  - Performance tests

- Operational product
  - Support analysis
  - Interaction logs
  - Field studies
  - Acceptance tests

References

- Alan Dix, Janet Finlay, Gregory Abowd and Russell Beale. (1998) Human Computer, Interaction (second edition), Prentice Hall, ISBN 0132398648 (new Edition announced for October 2003)
- Ben Shneiderman. (1998) Designing the User Interface, 3rd Ed., Addison Wesley; ISBN: 0201694972

- Discount Usability Engineering
  http://www.useit.com/papers/guerrilla_hci.html
- Heuristic Evaluation
  http://www.useit.com/papers/heuristic/