Vorlesung Mensch-Maschine-Interaktion

Evaluation

Ludwig-Maximilians-Universität München
LFE Medieninformatik
Heinrich Hußmann & Albrecht Schmidt
WS2003/2004
http://www.medien.informatik.uni-muenchen.de/

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

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)

Why Studies and Experiments?

- To measure more precisely the usability or other features
- Applicable mainly to
  - Functional prototypes
  - Testing an implementation
  - Quality monitoring of software products
- To compare solutions, e.g.
  - Users are quicker using version A than using version B
  - Users make 10% less errors when using version X than when using version Y
- To provide quantitative figures, e.g.
  - 90% of the users can complete the transaction using version Y in less than 3 minutes
  - On average users will be able to by a ticket using version A in less than 30 seconds

Designing the experiment

- Basic Scientific Method
  1. Form Hypothesis
  2. Collect data
  3. Analyze
  4. Accept/reject hypothesis
- Issues for user studies
  - System to test
  - Participants
  - Hypothesis
  - Variables
  - Experimental Methods
  - Statistical approach

Does computer science fit this traditional science approach?

Is it really possible to prove usability?

Procedure for user studies

- Set goals
- Design the experiment
- Schedule users
- For each user (typical example):
  - Inform the user and sign the consent form
  - Do a survey on demographics and questions of interest to the experiment
  - Give the participant instructions on the task – do not reveal the hypotheses
  - (optional) Make a training run - depends on the study
  - Perform the actual run and measure variables
  - (optional) do a survey on subjective measure
  - Be available for questions of participants or for their (informal) feedback
- Analyze the results

Participants (Subjects)

- How many participants do we need?
  - Depending on the project and the goals
  - Depending on the set-up
    - measuring the login-in time of remote users vs.
    - Doing a full video observation for a 1 hour task
  - Be pragmatic
  - Minimal size of about 10 participants
- Participants should be representative for the user group
  - Age, background, skills, experience, …
  - In most cases the other people on the team are NOT representative!
- How to recruit participants
  - Customer data base
  - Market research services
  - Volunteers (online, newspapers, etc) – this is risky because the people who will respond are often not representative

Selection of Participants

- Services offered that allow to get participants fitting a specific description
- Methods widely used in market research

- Example: Online Panel
  - For online questionnaires
  - Pool of users
  - Customer can specify the users that should take part

- How do companies get their subjects?
  - Incentive (money, prices, …)
  - Big set of questions when registering potential users, show examples from ComCult Online Panel

Variables

- Variables are manipulated and measured
  - Independent variables are manipulated
  - Dependent variables are measured
- The conditions of the experiment are set by independent variables
  - E.g. number of items in a list, text size, font, color
  - The number of different values used is called *level*
  - The number of experimental conditions is the product of the levels
  - E.g. font can be times or arial (2 levels), background can be blue, green, or white (3 levels). This results in 6 experimental conditions (times on blue, times, on green, …, arial on white)
- The dependent variables are the values that can be measured
  - Objective values: e.g. time to complete a task, number of errors, etc.
  - Subjective values: ease of use, preferred option
  - They should only be dependent on changes of the independent variables

Hypotheses

- Prediction of the result of an experiment
- Stating how a change in the independent variables will effect the measured dependent variables
- With the experiment it can be shown that the hypotheses is correct
- Usual approach
  - Stating a null-hypotheses (this predicts that there is not effect of the change in the independent variable on the measured variable)
  - Carrying out the experiment and using statistical measures to disprove the null-hypotheses
  - When a statistical test shows a significant difference it is probable that the effect is not random

Designing the experiment

- The experiment should be set up to be reproducible!
- Main factors
  - Participants
  - Independent variables
  - Hypotheses stated
- Approach
  - state the hypotheses – what do you want to proof
  - find the variables? Which are varied? which are measured?
  - Find participants – representative for the experiment
  - Fix the method to use (between-groups / within groups)

Experimental Method

- Within groups
  - Each user performs under all the different conditions
  - Important to randomize the order of the conditions for each participant
  - Problems
    - Learning may influence results
  - Advantages
    - The effect of differences between individuals are lessened
    - Fewer participants required
- Between groups (randomize)
  - One condition is selected for each participant
  - Each user performs only under one condition (avoids learning)
  - Careful selection of groups is essential
  - Drawback
    - Differences between individuals in different groups can play an important role (leads to large groups)
    - More user required
    - Usually harder to show significance

Statistical Tests

- See statistics text book (e.g. form psychology or medical tests)
- Software packages offer functions
- Test selected depends on
  - Distribution of the measured variables
  - The type of variables (continuous or discrete)
  - Experimental Method
- Example: Student's t-test
  - On the difference of means
  - Assumes a normal distribution
  - Functions available in spreadsheet software and statistics packages
- Example ANOVA
  - Analysis of Variance

- "significant difference"
  - Simplified: the probability that effect observed is random is less the 0.05

T-Test example in Excel

- TTEST(…)

- Parameters
  - Data row 1
  - Data row 2
  - Ends (1 or 2)
  - Type (paired, same variance, different variance)

User	Time M1	Time M2		
100	37	31		
101	44	38		
102	42	43		
103	56	37		
104	99	50		
105	33	30		
106	45	50		
107	49	36		
108	70	71		
109	63	56		
110	54	51		
111	61	46		
average	54,4167	44,9167		
t test (paired)			0,042	TTEST(B7:B18;C7:C18;2;1)
t test (un-paired)			0,137	TTEST(B7:B18;C7:C18;2;2)

Further Issues

- Consent form – get written consent from participants
  - Templates available
  - 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

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:_____

Example: Study on Text Input

- Is text input by keyboard really better than using T9 on a phone?
- Compare text input speed and errors made
  - Qwertz-keyboard on a notebook computer
  - T9 on a mobile phone
- Concentrate on test input only, ignore:
  - Time to setup / boot / initialize the device
  - Time to get into the application

Example: Study on Text Input (2)

- Participants
  - How many?
  - Skills
    - Computer user?
    - Phone/T9 users?
- Independent variables
  - Input method
  - Text to input
- Dependent variables
  - Time to input a text
  - Number of errors made

Example: Study on Text Input (3)

- Independent variables
  - Input method,
    - 2 levels: Keyboard and T9
  - Text to input
    - 1 level: text with about 10 words

- Experimental conditions
  - 2 conditions – T9 and Key
  - User 1,3,5,7,9 perform T9 than Key
  - User 2,4,6,8,10 perform Key than T9
  - Different texts in first and second run?
  - Particular phone model?
  - Completion time is measure (e.g. stop watch or application)
  - Number of error/corrections is observed

Example: Study on Text Input (4)

- Hypotheses
  - H-1: Input by keyboard is quicker than T9
  - H-2: fewer errors are made using keyboard input compared to T9

- Null-Hypotheses
  - Assumes no effect
  - H0-1: there is no difference in the input speed between keyboard and T9
  - H0-2: there is no difference in the number of errors made using a keyboard input compared to T9

- Experimental Method
  - Within groups
  - Randomized order of conditions

Example: Study on Text Input (5)

- Collect Data

User	Order	Time Cond1	Time Cond2	# Err Cond1	# Err Cond2
01	c1>c2	…	…	…	…
02	c2>c1	…	…	…	…
03	c1>c2	…	…	…	…

- Perform a statistical analysis
- … exercise on Monday.

Example: Study on Text Input (6)

- Fairness
  - Same conditions and procedure (e.g. light condition, interruptions, noise)
  - Specify procedure for exceptions (e.g. someone does not complete the test)
  - No bias
- Participants Consent
- Further Issues?
  - Ethical issues
  - Privacy

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:
O 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
O 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

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

Links

- http://www.studentenchallenge.de
  - Application development for Tablet PC

- http://www.imaginecup.com/
  - Categories
    - Software design
    - Rendering
    - Algorithm
    - Short Film

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