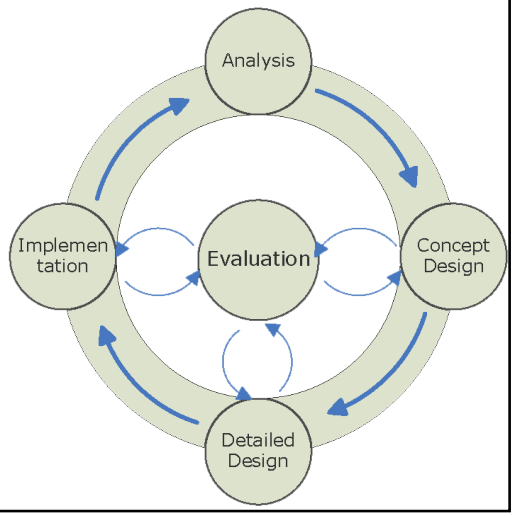
Human-Computer Interaction

Chapter 3

Human-Centered Aspects



Overview

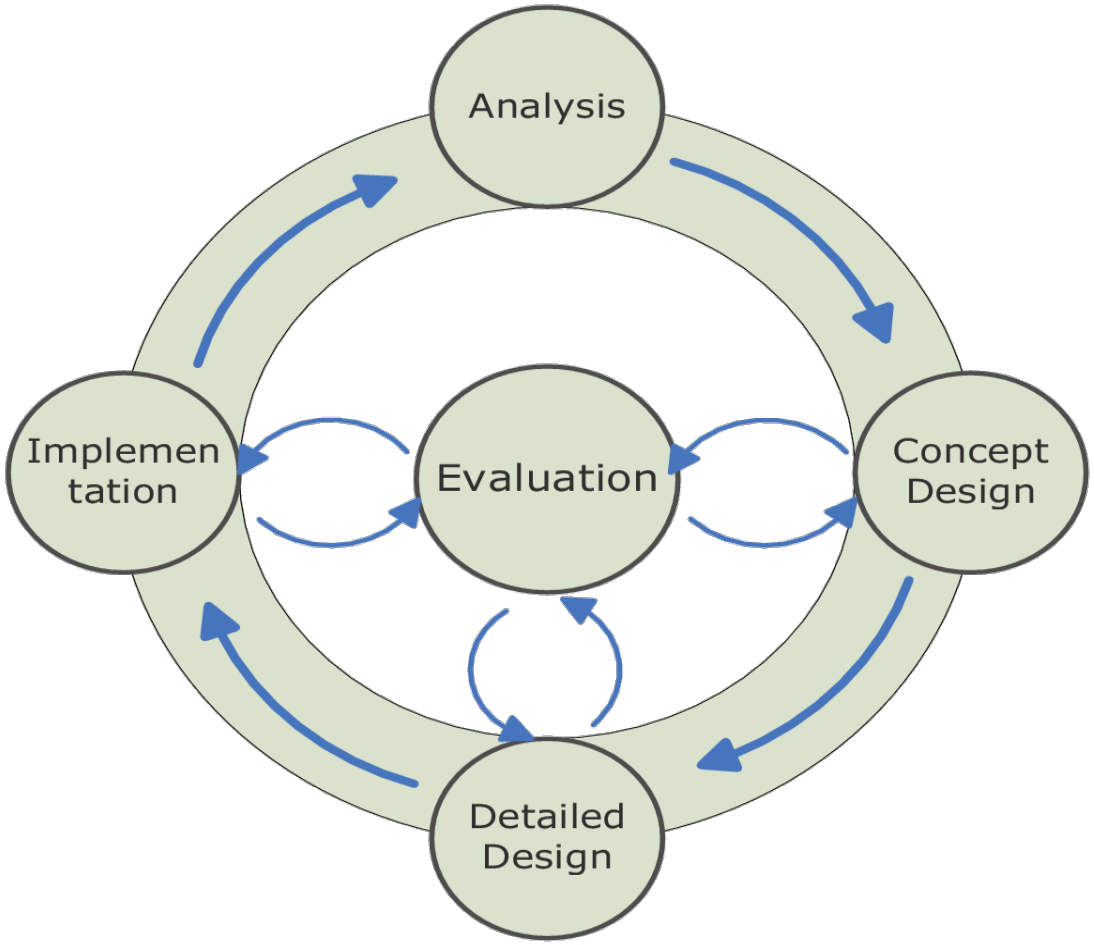
- 3.1 Usability in Human-Centered Design
- 3.2 User Aims and Requirements
- **3.3 Evaluation**
  - Evaluation cycle
  - Evaluation through expert analysis
  - Evaluation through user participation
  - Designing experiments
  - Classification of evaluation techniques

Evaluation

- Role of Evaluation
  - Need to assess designs
  - Need to test systems
  - Ensure that systems behave as expected
  - Ensure that systems meet user requirements
- Goals of Evaluation
  - Assess extent and accessibility of system's functionality
  - Assess users' experience of interaction
  - Identify any specific problems with the system
- When?
  - Throughout the design life cycle!
  - Results should feed back into design modifications

Evaluation

- Evaluation Cycle / User-Centered Design Cycle



Evaluation

- Evaluation Methods
  - Formative vs. summative
    - Evaluation during or after design phase
  - Analytical vs. empirical
    - Analytical methods:
      Evaluation through expert analysis
      - Cognitive walkthrough
      - Heuristic evaluation
      - Early concept evaluation
    - Empirical methods:
      Evaluation through user participation
      - Controlled experiments
      - Observational evaluation
      - Query techniques: Interviews, Questionnaires

Analytical Evaluation

- Evaluation through Expert Analysis
  - Basic intention
    - Identify areas that are likely to cause difficulties
    - Reason 1: Violation of known cognitive principles
    - Reason 2: Ignore accepted empirical results
  - Methods can be used at any stage of design process
  - Properties
    - Often done by usability experts evaluating the preliminary design
    - Relatively cheap (they do not require user involvement)
    - But: they do not assess actual use of the system

Analytical Evaluation

- Cognitive Walkthrough
  - Walkthroughs require detailed review of a sequence of actions
  - Origin: code walkthrough (SE)
    - Sequence = segment of program code
    - Sequence is stepped through by reviewers to check certain characteristics (e.g., code styling)
  - Cognitive walkthrough
    - Sequence = steps that an interface will require a user to perform in order to accomplish a given task
    - Evaluators check it for potential usability problems
    - Main focus usually on how easy system is to learn

Analytical Evaluation

- Cognitive Walkthrough (cont.)
  - Things needed for walkthrough
    - Specification or prototype of the system
    - Description of the task the user is to perform
    - Complete list of actions needed to complete the task
    - Indication of who users are (experience, knowledge)
  - For each action and each step in the action sequence answer the following questions
    - Is effect of action the same as the user's goal at that point?
    - Will users see that the action is available?
    - Once users have found correct action: Will they know it is the one they need?
    - After the action is taken: Will users understand the feedback they get?

Analytical Evaluation

- Heuristic Evaluation
  - Heuristic
    - Meaning: guideline, general principle
    - Can guide a design decision
    - Can be used to critique a decision
  - Heuristic evaluation
    - Method for structuring a critique of a system
    - Developed by Nielson and Molich
    - Flexible and relatively cheap approach
      - Uses a set of relatively simple and general heuristics
      - Useful for evaluating early designs
      - Can also be used on prototypes, storyboards, fully functioning systems

Analytical Evaluation

- Heuristic Evaluation (cont.)
  - General idea
    - Several evaluators independently critique a system
    - Goal: come up with potential usability problems
    - Number of evaluators needed [Nielson]
      - Between three and five is sufficient
      - Five usually result in ~75% of the overall usability problems being discovered
  - Evaluators assess the severity of each problem (rating on a scale of 0-4)
    - 0 = no problem at all
    - 1 = cosmetic problem only (need not to be fixed)
    - 2 = minor usability problem (fixing has low priority)
    - 3 = major usability problem (important to fix)
    - 4 = usability catastrophe (imperative to fix)

Analytical Evaluation

- Heuristic Evaluation (cont.)
  - Set of 10 heuristics [Nielson]
    1. Visibility of system status
    2. Match between system and real world
    3. User control and freedom (e.g. support undo + redo)
    4. Consistency and standards
    5. Error prevention
    6. Recognition rather than recall
    7. Flexibility and efficiency of use
    8. Aesthetic and minimalistic design
    9. Help users recognize, diagnose and recover from errors
    10. Help and documentation

Analytical Evaluation

- Early concept evaluations
  - A way to ask users whether they like a future design and would use it
  - Most common technique: Focus Groups
    - Special form of group interviewing
    - 8 – 12 participants
    - Interaction within the group based on topics supplied by researcher
    - Not natural but organized events
    - To gain large amount of information in short time period
    - Get users' attitudes, feelings, beliefs, experiences, and reactions

Analytical Evaluation

- Early concept evaluations (cont.)
  - Guideline: Focus Group Procedure
    - Identify purpose and audience
    - Develop screener document:
      document helps you select a well-rounded group;
      screener = questionnaire (≤ 5min per participant)
    - Recruit participants
    - Sign up and confirm participants
    - Develop discussion guide:
      first general questions; then more specific probes
    - Plan session logistics
    - Run focus group
    - Debrief informally (immediately after session)
    - Analyze notes and/or tapes
    - Write focus group report

Analytical Evaluation

- Early concept evaluations (cont.)
  - Guideline: Running a Focus Group
    - Introduction and orientation
    - General issues
    - Issues specific to project
    - Viewing of and reaction to product
    - Appropriateness of content for the target audience
    - Discussion of product usefulness
    - Conclusions
    - Thank participants

Empirical Evaluation

- Analytical evaluation
  - Useful for filtering and refining design, but no replacement for usability testing with users!
- Empirical evaluation: user participation in evaluation
  - Usually occurs in later stages
  - At least a working prototype needed (may range from a simulation of interactive capabilities to a fully implemented system)
- Styles of evaluation
  - Laboratory studies
  - Field studies

Empirical Evaluation

- Laboratory studies
  - Users are taken out of normal work environment
  - Users take part in controlled tests (interruption-free environment)
  - Unnatural situation ➪ lack of context
  - But: possibility of context control / manipulation
- Field studies
  - Takes evaluator out into user's environment
  - Context retained, but high level of ambient noise
  - Interactions may be observed that would have been missed in laboratory study
- Both: participants likely to be influenced by presence of analyst and/or recording equipment

Empirical Evaluation

- Experimental evaluation: Method
  - Controlled experiment
  - Provides empirical evidence to support a hypothesis
  - Can be used to study a wide range of different issues at different levels of detail
  - Any experiment has same basic form
    - Evaluator chooses a hypothesis to test
    - Hypothesis need to be determined by measuring some attributes of participant behavior
    - Different experimental conditions differ only in the values of certain controlled variables
  - Factors important for reliability include
    - Participants chosen
    - Variables tested and manipulated
    - Hypothesis tested

Empirical Evaluation

- Experimental evaluation: Participants
  - Choice of participants
    - Vital to success of any experiment
    - Participants should match expected user population as closely as possible
  - Sample size
    - Must be large enough to be considered to be representative of the population
    - Usability testing with a single participant: will find about 1/3 of usability problems
    - Little to be gained from testing with more than five
    - Controlled experiment + statistical analysis: at least twice this number recommended

Empirical Evaluation

- Experimental evaluation: Variables
  - Experiments manipulate and measure variables under controlled conditions
  - Main type of variables: independent, dependent
  - Independent variables (ivar)
    - Manipulated to produce different conditions for comparison
    - Examples: interface style, number of menu items
    - Complex experiments: more than one ivar
  - Dependent variables (dvar)
    - Dvars must be measurable, affected by ivar(s), and (as far as possible) unaffected by other factors
    - Value is dependent on changes made to ivar(s)
    - Examples: Time to complete task, number of errors, user preference

Empirical Evaluation

- Experimental evaluation: Hypothesis
  - Hypothesis = prediction of outcome of experiment
  - Framed in terms of ivars and dvars
  - Prediction: variation in ivar ⇒ difference in dvar
  - Aim of experiment: show that prediction is correct
    - Done by disproving the null hypothesis “no difference in dvar between levels of ivar”
  - Statistical analysis
    - Choice of analysis method depends on type of data and questions to be answered
    - Produces values that can be compared with various levels of significance
    - If a result is significant, the differences measured would not have occurred by chance
      ⇒ null hypothesis is incorrect

More information and examples on statistical analysis: HCI book of A. Dix !

Empirical Evaluation

- Experimental evaluation: Experimental design
  - First step: choose hypothesis
  - Next step: choose experimental method
    - Between-subjects (randomized design)
      - Each participant is assigned to different condition
      - At least two conditions: experimental (ivar manipulated) and control (ivar not manipulated)
      - Control serves to ensure that it is the manipulation that is responsible for any measured differences
      - Significant variation between groups can negate results
      - Advantage: no learning effects
    - Within-subjects (repeated measures)
      - Each user performs under each different condition
      - Can suffer from transfer of learning effects
        ⇨ Can be lessened if order is varied between users
      - Advantages: less costly (fewer users required), less chance of effects from variation between participants

Empirical Evaluation

- Observational evaluation
  - Evaluator watches and records users‘ actions
  - Think Aloud
    - User observed performing task
    - User asked to describe what he is doing and why, what he thinks is happening, etc.
    - Advantages
      - Simplicity - requires little expertise
      - Can provide useful insight
      - Can show how system is actually used
    - Disadvantages
      - Subjective
      - Selective
      - Act of describing may alter task performance
    - Usefulness largely dependent on effectiveness of recording method (protocol)

Empirical Evaluation

- Observational evaluation (cont.)
  - Protocol analysis
    - Paper and pencil
      - Cheap, limited to analyst's writing speed
    - Audio
      - Good for think aloud, difficult to match with other protocols
    - Video
      - Accurate and realistic, needs special equipment, choosing camera position is difficult, obtrusive
    - Computer logging
      - Automatic, cheep, unobtrusive, large amounts of data (at keystroke level), difficult to analyze
    - Mixed use in practice
    - Audio/video transcription difficult and requires skill

Empirical Evaluation

- Designing experiments
  - With whom?
    - Find balance in user profile
      - Genders
      - Age range
      - Level of experience with particular technology
      - Cultural diversity
      - Personality differences
      - …
  - How?
    - Design requires expertise from several domains
      - Consult evaluation and domain experts
    - Tasks have to be realistic, motivating, and engaging
    - Recommended approach
      - Let users explore interface freely
      - Using their own data

Empirical Evaluation

- Designing experiments (cont.)
  - When and where
    - Estimate the time for preparation
    - Find participants
    - Choose an optimal location
      - Lab
      - Natural environment
      - Simulated environment close to real conditions
    - Perform a pilot test !!
      - Measure how much time each part and the whole experiment will take
      - Essential to discover the pitfalls that could happen during the real test

Empirical Evaluation

- Designing experiments (cont.)
  - How to avoid being biased
    - Maybe invite an outside experiment facilitator
      - But: some details of the experiment could be forgotten
    - Split roles
      - Experiment facilitator, technical support with prototype, observer
      - Use separate experiment protocol for each person
  - Ensure privacy of users
    - Disassociate personal information and data
    - Video and audio recordings should remain confidential
    - Use a special consent form
      - Explain aim of user study
      - Asks permission for audio/video recordings
      - Ensure participants that personal details will be used only for the purpose stated

Empirical Evaluation

- Designing experiments (cont.)
  - When to stop evaluating
    - Every evaluation test will reveal some area in which improvements can be made
    - There is never a time when interface is perfect
    - Budget / time constraints usually determine stop
    - „Quick and dirty“- techniques are sometimes a reasonable alternative
    - Expert analysis is a reasonable first step
      - Will not discover all problems
      - Important to let potential users explore the interactive design

Classification of Evaluation Techniques

- Factors distinguishing evaluation techniques
  - Stage in the cycle at which evaluation is carried out
    - Design vs. implementation
  - Style of evaluation
    - Laboratory vs. field studies
  - Level of subjectivity or objectivity of used technique
  - Type of measures provided
    - Quantitative vs. qualitative measurement
  - Level of information provided
  - Immediacy of response
  - Level of interference implied
  - Resources required

Classification of Evaluation Techniques

	Cognitive walkthrough	Heuristic evaluation
Stage	Throughout	Throughout
Style	Lab	Lab
Objective?	No	No
Measure	Qualitative	Qualitative
Information	Low level	High level
Immediacy	---	---
Intrusive?	No	No
Time	Medium	Low
Equipment	Low	Low
Expertise	High	Medium

Classification of Evaluation Techniques

	Experiment	Interviews	Questionnaire
Stage	Throughout	Throughout	Throughout
Style	Lab	Lab/field	Lab/field
Objective?	Yes	No	No
Measure	Quantitative	Qualitative / Quantitative	Qualitative / Quantitative
Information	Low/high level	High level	High level
Immediacy	Yes	No	No
Intrusive?	Yes	No	No
Time	High	Low	Low
Equipment	Medium/high	Low	Low
Expertise	Medium	Low	Low

Classification of Evaluation Techniques

	Think aloud [1]	Protocol analysis [2]
Stage	Implementation	Implementation
Style	Lab/field	Lab/field
Objective?	No	No
Measure	Qualitative	Qualitative
Information	High/low level	High/low level
Immediacy	Yes	Yes
Intrusive?	Yes	Yes [3]
Time	High	High
Equipment	Low	High
Expertise	Medium	High

1 Assuming simple paper + pencil record

2 Including video, audio, and system recording

3 Except system logs