
Human Computer Interaction

Evaluation Techniques

Lecture # 11

Imran Siddiqi
imran.siddiqi@gmail.com

Heuristic Evaluation

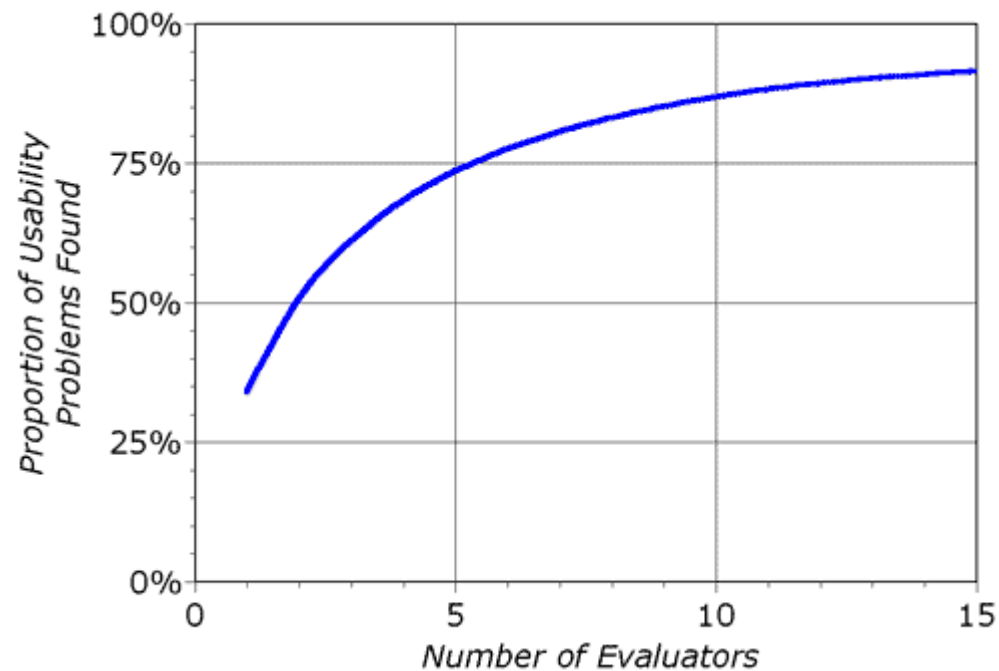


Heuristic Evaluation

- Heuristics
 - Guidelines, General Principle, Rule of thumbs
- Heuristic Evaluation
 - Critique of a system using a set of simple and general heuristics
 - Developed by Nielsen and Molich

Heuristic Evaluation

- Several evaluators independently critique a system for potential problems
- Nielsen's Experience: 3-5 evaluators



Heuristic Evaluation

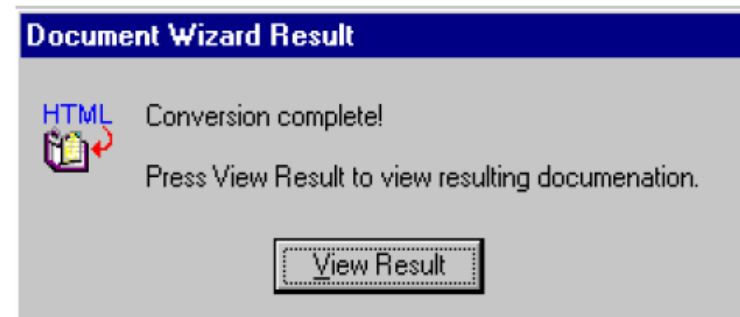
- Neilson's 10 heuristics
- Can be supplemented by domain specific heuristics if required
- Evaluators
 - Asses the system and note the violations of these heuristics
 - Also asses the severity of the problem

Heuristic Evaluation

- Problem Severity - Rating on a scale 0-4
 - 0 = I do not agree that this is a usability problem
 - 1 = Cosmetic Problem only, need not to be fixed unless extra time is available
 - 2 = Minor Usability Problem, Low priority fixation
 - 3 = Major Usability Problem, High priority fixation
 - 4 = Usability Catastrophe, Imperative to fix

Nielsen's 10 Heuristics

- 1. Visibility of System Status
 - Always keep user informed about what is going on?
 - Appropriate feedback – How long/ how much completed
 - Cursor Change, Progress Bar
 - Do not OVERDO it!



Nielsen's 10 Heuristics

■ 1. Visibility of System Status (Contd...)

■ Response Time

- < 0.1 secs – Instantaneous
- 0.1 – 1 secs – User notices, but no feedback needed
- 1 – 5 secs – Display busy cursor
- > 1 - 5 secs – Display progress bar



Nielsen's 10 Heuristics

- 2. Match between system and the real world
 - *System should speak user's language not system oriented terms*
 - *Words, phrases and concepts – familiar to the user*



Nielsen's 10 Heuristics

- 3. User control and freedom
 - *Users often make mistake and will need a clearly marked "emergency exit" to leave the unwanted state*
 - *Support undo and redo*
 - *Long operations should be cancelable*
 - *All dialogs should have a cancel button*

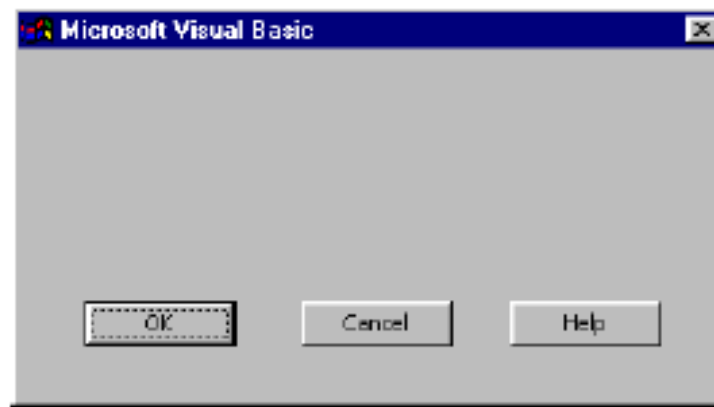
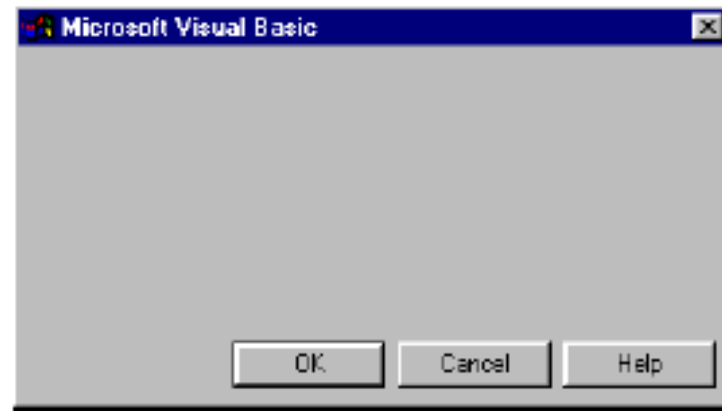
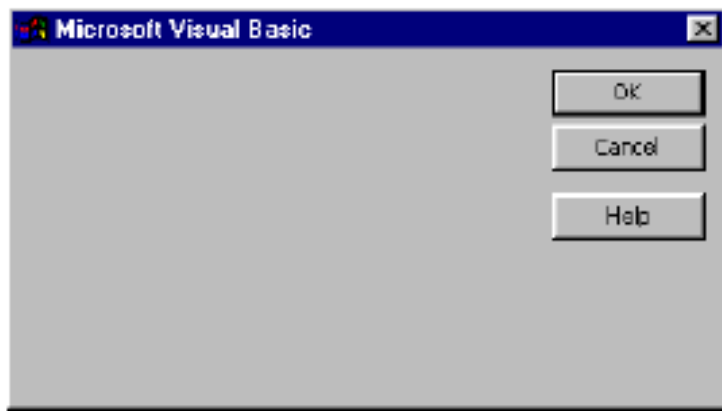


Nielsen's 10 Heuristics

- 4. Consistency and Standards
 - *Principle of least surprise*
 - *Similar things should look and act similar*
 - *Different things should look different*
 - *Properties*
 - *Terminologies, Size, Location, Color, ...*
 - *Follow standards*

Nielsen's 10 Heuristics

■ 4. Consistency and Standards



Nielsen's 10 Heuristics

■ 5. Error Prevention

- *Make it difficult to make errors*
- *Better than good error messages is careful design which prevents a problem from occurring in the first place*
- Selection is less error-prone than typing
 - Don't overdo it



Nielsen's 10 Heuristics

■ 5. Error Prevention (Contd...)

■ *Disable illegal commands*

- Copy is impossible if nothing is selected – then the command should be disabled (“grayed out”) so that it simply can’t be selected in the first place.

■ *Avoid Modes*

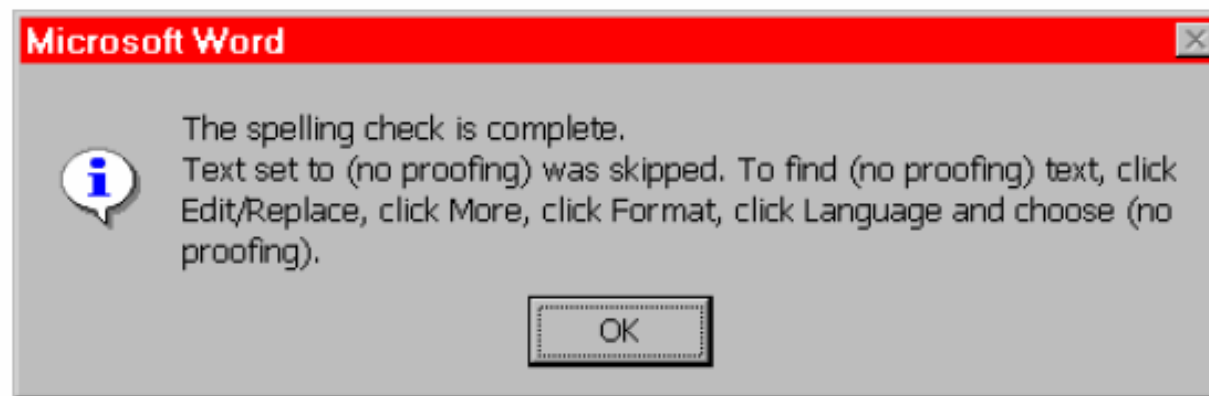
- If cant avoid – make it visible

Nielsen's 10 Heuristics

- 6. Recognition rather than recall
 - *Minimize the user's memory load by making objects, actions, and options visible*
 - *Instructions for use of the system should be visible or easily retrievable whenever appropriate*
 - *Use menus not command language*
 - *Use combo boxes not text boxes*

Nielsen's 10 Heuristics

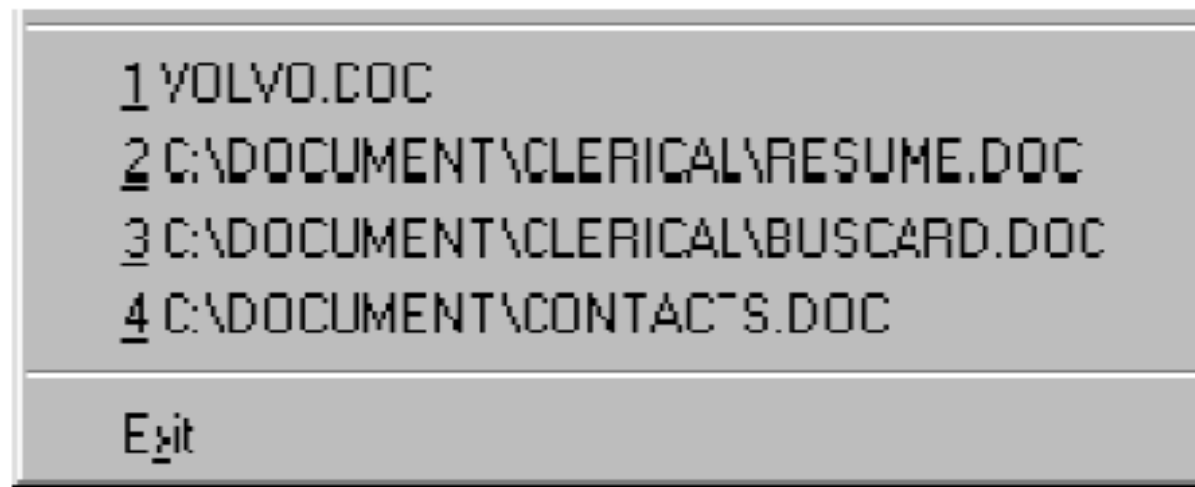
- 6. Recognition rather than recall
 - *Over reliance on user's memory*
 - *Modal dialog box, so the user can't start following its instructions until after clicking OK*
 - *Clicking OK the instructions vanish from the screen, and the user is left to struggle to remember them*



Nielsen's 10 Heuristics

■ 7. Flexibility and Efficiency

- *Accelerators to speed up the interaction for expert user*
- *Can cater to both inexperienced and experienced users*
 - Keyboard accelerators, Command abbreviations, Bookmark, History

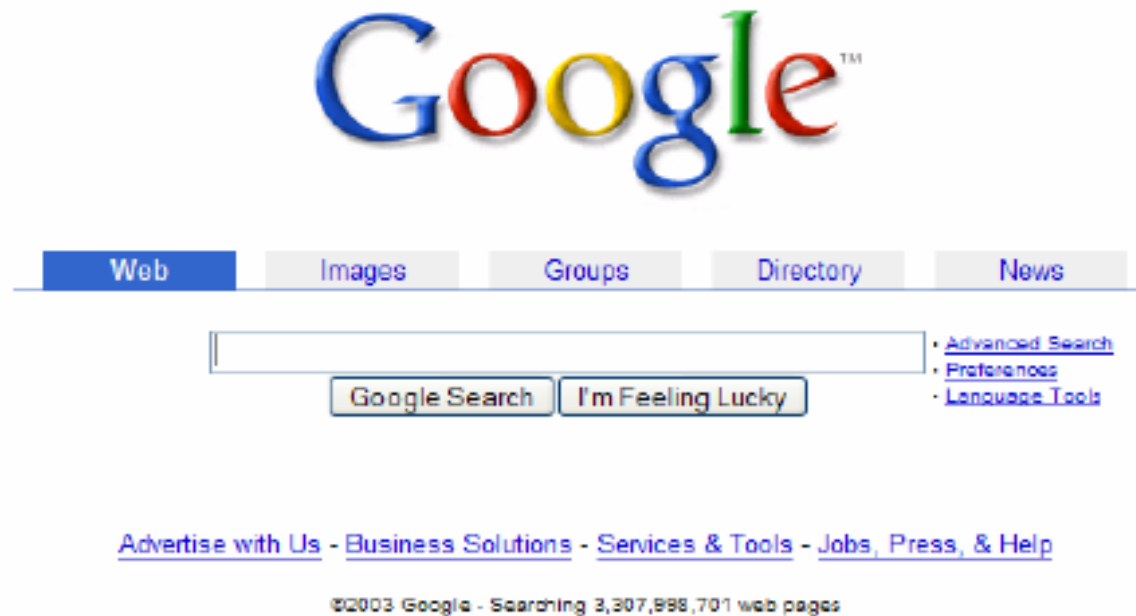


Nielsen's 10 Heuristics

- 8. Aesthetic and minimalist design
 - *Less is More*
 - *Do not use information which is irrelevant or rarely needed*
 - *Every extra unit of info competes with the relevant units of info – diminishes their relative visibility*
 - *Omit extraneous info, graphics, features*

Nielsen's 10 Heuristics

■ 8. Aesthetic and minimalist design

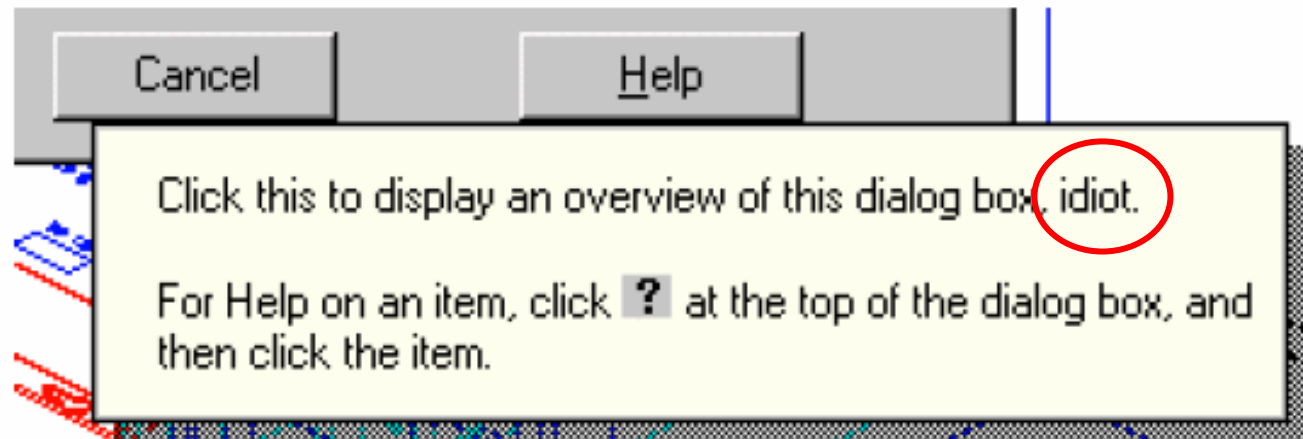


Nielsen's 10 Heuristics

- 9. Help users recognize, diagnose, and recover from errors
 - *If you cant prevent errors – give a good error message*
 - *Should be in plain language (No codes)*
 - *Be precise – Restate user's input*
 - Not “Cannot open file”, but “Cannot open file named paper.doc”
 - *Give constructive help*
 - Why error occurred and how to fix it
 - Be polite and non-blaming
 - Not “fatal error”, not “illegal”

Nielsen's 10 Heuristics

- 9. Help users recognize, diagnose, and recover from errors (Contd ...)
 - The tooltip shown comes from a production version of AutoCad!
 - Inserted by a programmer as a joke, but somehow never removed before release.



Nielsen's 10 Heuristics



- 10. Help and Documentation
 - *Users don't read manuals*
 - Prefer to spend time working toward their task goals, not learning about your system
 - *But manuals and online help are vital*
 - Usually when user is frustrated or in crisis
 - *Help should be:*
 - Searchable, Concrete and Short

Evaluation through User Participation

Actual User not designer or expert
Later stage of development

Styles of Evaluation

- Lab Studies

- *Users taken out of their normal work environment to take part in controlled test*
- *Lab could be just a “Quiet Room”*

- Field Studies

- *Take evaluator to user’s work environment*

Lab Studies

- Advantages:

- Specialist equipment available – Audio Visual Recording etc.
- Uninterrupted environment

- Disadvantages:

- Lack of context – Wall papers, cabinets, disturbances (Not close to reality)
- Difficult to observe several users cooperating

- Appropriate

- If system location is dangerous or remote
- War environment and space station

Field Studies

- Advantages:

- Natural environment
- Might observe interactions between system and individual that may be missed in a lab study
- Longitudinal studies possible Days or Months

- Disadvantages:

- Distractions (Phones, visitors)
- Noise

Empirical Methods – Experimental Evaluation

- Evaluating a design in controlled experiment
- Evaluator chooses hypothesis to be tested
- A number of experimental conditions are considered which differ only in the value of some controlled variable
- Changes in behavioural measure are attributed to different conditions

Experimental Factors

- Participants

- Who – representative, sufficient sample

- Variables

- Things to modify and measure

- Hypothesis

- What you'd like to show

- Experimental design

- How you are going to do it

Participants

- Participants should be chosen to match expected user population as close as possible
 - Ideally end user but not always possible
 - At least same age, educational background, experience, etc...
- Similar expertise and domain knowledge
 - Is a engineering background representative of general population?
- Sample size
 - Large sample size considered representative of population

Variables

- Independent variable (IV)
 - Characteristic changed to produce different conditions
 - E.g. interface style, number of menu items, icon design
- Dependent variable (DV)
 - Characteristics that can be measured in the experiment
 - Their value is dependent on changes made to Independent variables (IV)
 - E.g. time taken to perform a task, number of errors etc;

Hypothesis

- Prediction of outcome of an experiment
- Aim of experiment is to show that this predictions are correct
 - Framed in terms of IV and DV
 - E.g. “error rate will increase as font size decreases”
- Null hypothesis:
 - States no difference between conditions
 - Aim is to disprove this
 - E.g. null hyp. = “no change with font size”

Experimental Design

- First phase is to choose hypothesis
 - What are you trying to prove
- Clarify independent and dependent variables
- Consider your participants
- Choose experiment method
 - Within- Subjects
 - Between- Subjects

Experimental Design

- Within subjects design
 - Each subject performs experiment under each condition
 - Transfer of learning possible
 - Less costly and less likely to suffer from user variation
- Between subjects design
 - Each subject performs under only one condition
 - No transfer of learning
 - More users required
 - Variation can bias results

Observational Methods

(Observe users interact with a system)

Think Aloud
Cooperative evaluation
Post-task walkthroughs

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 use
- Disadvantages
 - Performance may not be natural in the presence of observer
 - Act of describing may alter task performance



Cooperative Evaluation

- Variation on think aloud
- User collaborates in evaluation
- Evaluator asks questions - “Why?” “What if?”
- Additional advantages
 - Less constrained and easier to use
 - User is encouraged to criticize system
 - Clarification possible

Methods for recording user actions

- Paper and pencil

- Cheap, limited to writing speed



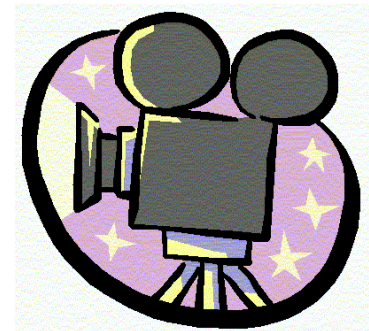
- Audio

- Good for Think Aloud



- Video – accurate and realistic

- Camera position to show sufficient details and user is difficult
 - May ask user not to move but that is not normal behaviour



Methods for recording user actions

- Computer logging – automatically done
 - Large amounts of data difficult to analyze
- User notebooks
 - Coarse and subjective
 - Good for longitudinal studies
- Mixed use in practice
 - Audio/video transcription with paper record of significant events



Post-Task Walkthroughs

- Observational transcript (written or recorded) played back to participant for comment
- Participant's comments
- Immediately → Fresh in mind
- Delayed → Evaluator has time to identify questions
- Necessary in cases where think aloud is not possible



Query Techniques

Interviews
Questionnaires

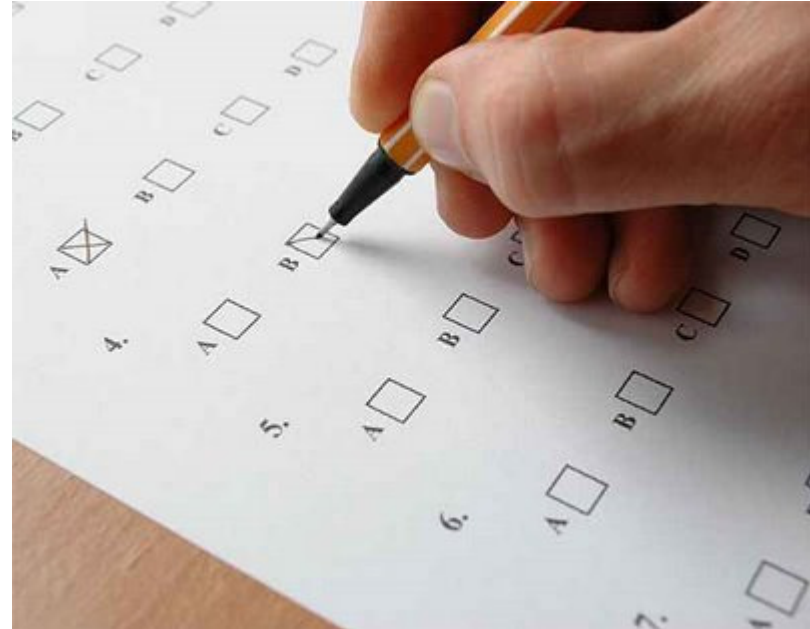
Interviews

- Analyst questions user on one-to-one basis
 - Usually based on prepared questions
 - Informal, subjective and relatively cheap
- Advantages
 - Can be varied to suit context
 - Issues can be explored more fully
- Disadvantages
 - Subjective and Time Consuming



Questionnaires

- Set of fixed questions given to users
- Quick and reaches large user group
- But less flexible



Questionnaires

- Need careful design
 - What information is required?
 - How are answers to be analyzed?
- Styles of question
 - General → “Establish background of user” Age, Sex, Job
 - Open-ended → Can you suggest any improvement?
 - Scalar → 1 2 3 4 5 Is it easy to recover from mistakes?
 - Multi-choice → How do you get help? Different choices
 - Ranked → Rank usefulness of methods to issue command

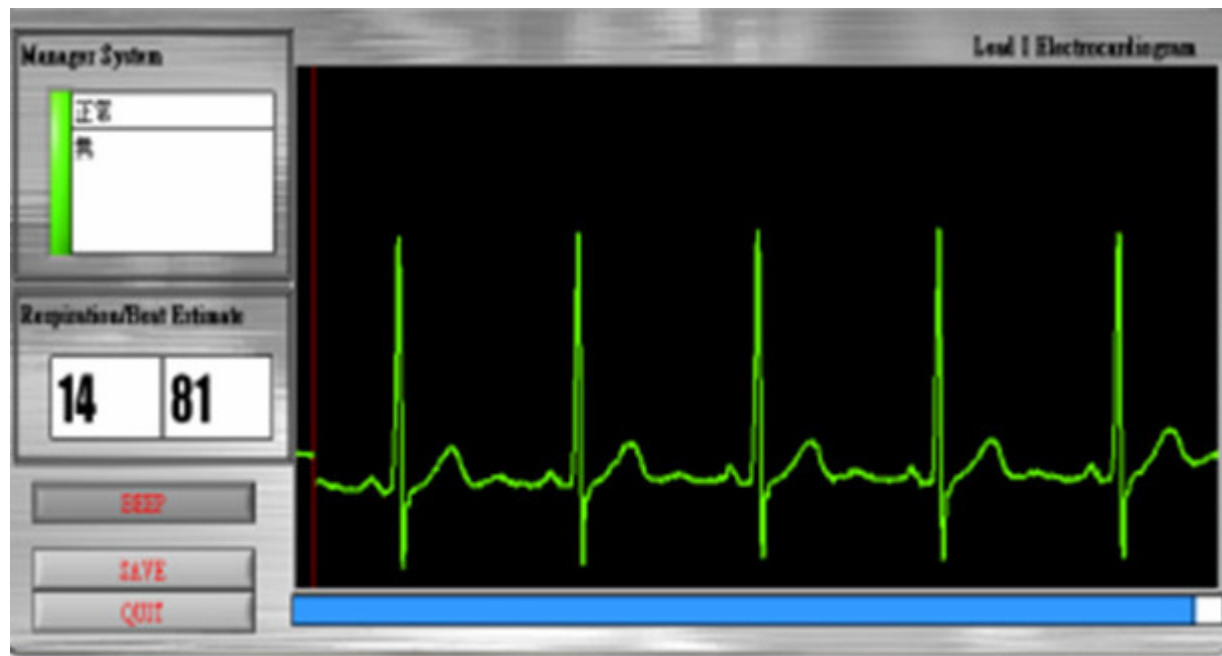
Questionnaires

- Comprehensive?
 - Test on 4-5 users before sending out
- Distribution
 - Respondents representative of user population
- Return rate
 - Very low – 25% to 30%

Physiological Methods

Eye Tracking

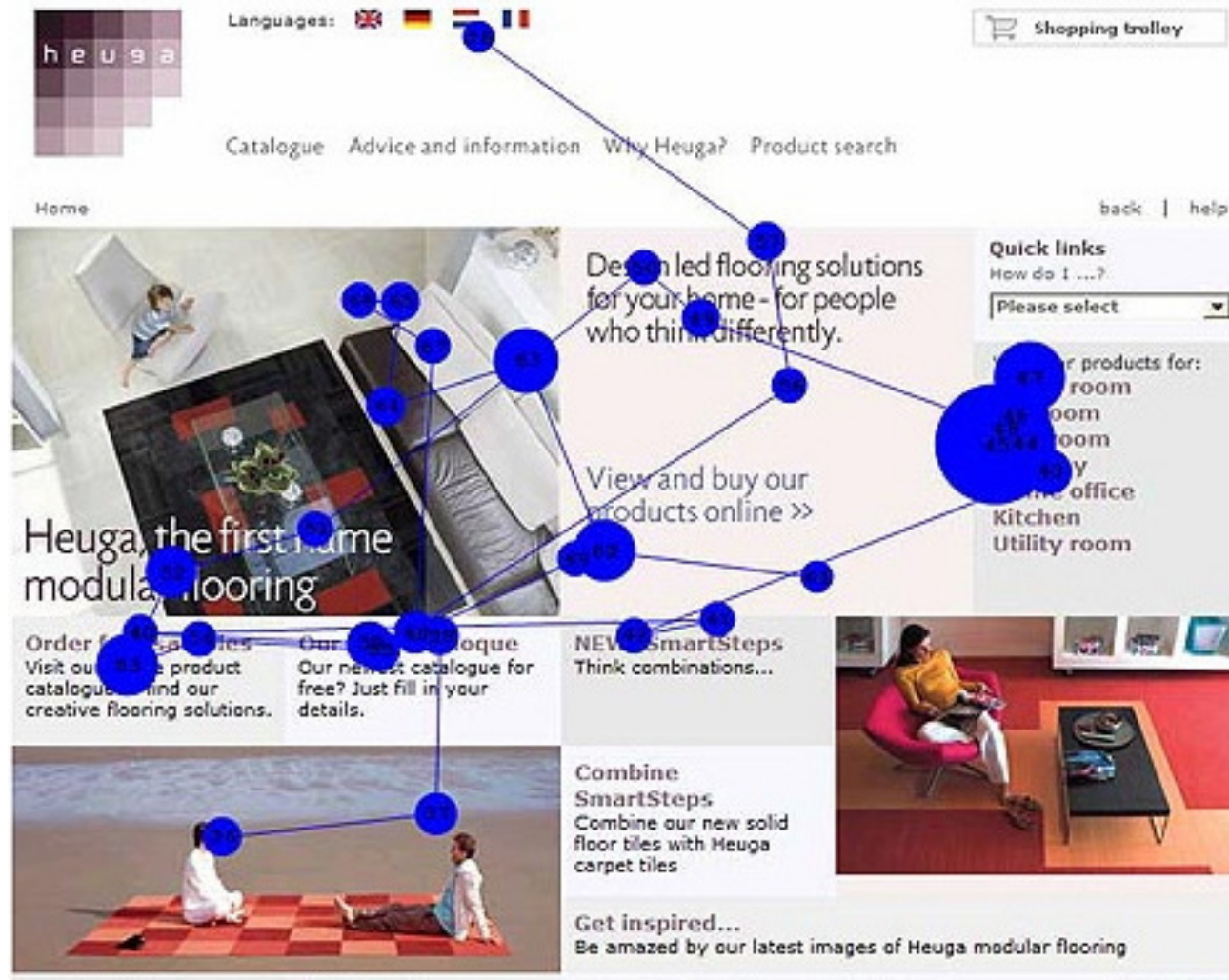
Physiological Measurements



Eye Tracking

- Head or desk mounted equipment tracks the position of the eye
- Eye movement reflects the amount of cognitive processing a display requires
- Measurements include
 - **Fixations**: Eye maintains stable position. Number and duration indicate level of difficulty with display
 - **Saccades**: Rapid eye movement from one point of interest to another
 - **Scan paths**: Moving straight to a target with a short fixation at the target is optimal

Eye Tracking



Physiological Measurements

- These may help determine a user's reaction to an interface
- Measurements include:
 - Heart activity, including blood pressure, pulse etc.
 - Activity of sweat glands
 - Activity in muscle
 - Activity in brain
- Difficulty in interpreting these physiological responses - more research needed

Choosing an Evaluation Method

- Design vs. Implementation
 - At what stage in design process the evaluation is required
 - Evaluation of design – Should be quick and cheap
 - Design Experts only
 - Evaluation of Implementation
 - Comprehensive – User Participation

Lab vs. Field Studies

- Lab Studies

- Controlled observation losing naturalness of user environment

- Field Studies

- Retain user environment but no control over activity

- Combine the two

- Lab studies in early stages and field studies with implementation

Subjective vs. Objective

- Subjective Methods

- Cognitive Walkthrough, Think Aloud – Rely heavily on the knowledge and interpretation of expert
- Evaluator bias – could be a problem

- Objective Methods

- Controlled Experiments – Avoid bias and produce comparable results

Qualitative vs. Quantitative Measures

- Related to subjectivity and objectivity
- Numeric Data
 - Statistical Analysis
- Non-numeric Data
 - Difficult to analyze
- Can quantify qualitative information
 - User response on a scale

Information Provided

- Could be low level or high level
- High level
 - Is the system usable?
 - Interviews
- Low level
 - What is the minimum font size?
 - Controlled Experiments

Immediacy of Response

- Think aloud
 - Record user behavior at the time of interaction itself
 - Measurement can alter the way user actually works
- Post-Task Walkthrough
 - Rely on user's recollection of events
 - Recall could be incomplete or biased

Intrusiveness

- Immediate measurements – influence the way a user behaves
- Automatic System logging
 - Limited Information

Resources

- Availability of resources
 - Equipment, Time, Money, Participants, Evaluators
 - Video recording not possible w/o a video camera
 - May not be possible to test the system in its intended environment
 - Space station

References

- Chapter 9 - Human Computer Interaction by Dix et al.
- Usability Engineering – Rob Miller, MIT
- HCI & Interface Design – Jacek Gwidza

