PV182
Human Computer Interaction

Lecture 5

Evaluating Interfaces

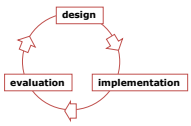
Fotis Liarokapis
liarokap@fi.muni.cz

15th October 2018

Evaluation Methods

Importance

- Tied to the usability engineering lifecycle
- Pre-design
  - Investing in new expensive system requires proof of viability
- Initial design stages
  - Develop and evaluate initial design ideas with the user



Importance .

- Iterative design
  - Does system behavior match the user's task requirements?
  - Are there specific problems with the design?
  - What solutions work?
- Acceptance testing
  - Verify that system meets expected user performance criteria
  - 80% of 1st time customers will take 1-3 minutes to withdraw $50 from the automatic teller

Overview

- Evaluation tests the usability, functionality and acceptability of an interactive system
- Evaluation may take place
  - In the laboratory
  - In the field
- Some approaches are based on expert evaluation
  - Analytic methods
  - Review methods
  - Model-based methods
- Some approaches involve users
  - Experimental methods
  - Observational methods
  - Query methods
- An evaluation method must be chosen carefully and must be suitable for the job

Naturalistic Approach

- Observation occurs in realistic setting
  - Real life
- Problems
  - Hard to arrange and do
  - Time consuming
  - May not generalize

Experimental Approach

- Experimenter controls all environmental factors
  - Study relations by manipulating independent variables
  - Observe effect on one or more dependent variables
  - Nothing else changes
- There is no difference in user performance (time and error rate) when selecting an item from a pull down or a pull right menu of 4 items

Validity

- External validity
  - Confidence that results applies to real situations
  - Usually good in natural settings
- Internal validity
  - Confidence in our explanation of experimental results
  - Usually good in experimental settings
- Trade-off: Natural vs Experimental
  - Precision and direct control over experimental design versus
  - Desire for maximum generalizability in real life situations

Usability Engineering Approach

- Observe people using systems in simulated settings
  - People brought in to artificial setting that simulates aspects of real world setting
  - People given specific tasks to do
  - Observations / measures made as people do their tasks
  - Look for problem areas / successes
  - Good for uncovering 'big effects'

Usability Engineering Approach .

- Is the test result relevant to the usability of real products in real use outside of lab?
- Problems
  - Non-typical users tested
  - Non-typical tasks
  - Different physical environment
  - Different social context
    - motivation towards experimenter vs motivation towards boss
- Partial Solution
  - Use real users
  - Task-centered system design tasks
  - Environment similar to real situation

Usability Engineering Approach ..

- How many users should you observe?
  - Observing many users is expensive
  - But individual differences matter
    - best user 10x faster than slowest
    - best 25% of users ~2x faster than slowest 25%

- Partial solution
  - Reasonable number of users tested
  - Reasonable range of users
  - Big problems usually detected with handful of users
  - Small problems / fine measures need many users

Discount Usability Evaluation

- Low cost methods to gather usability problems
  - Approximate: capture most large and many minor problems
- Qualitative:
  - Observe user interactions
  - Gather user explanations and opinions
  - Produces a description, usually in non-numeric terms
  - Anecdotes, transcripts, problem areas, critical incidents…
- Quantitative
  - Count, log, measure something of interest in user actions
  - Speed, error rate, counts of activities, etc

Discount Usability Evaluation .

- Methods
  - Inspection/cognitive walkthrough
  - Extracting the conceptual model
  - Direct observation
    - Think-aloud
    - Constructive interaction
- Query techniques
  - Interviews and questionnaires
- Continuous evaluation
  - User feedback and field studies

Inspection

- Designer tries the system (or prototype)
  - Does the system "feel right"?
  - Benefits
    - Can catch some major problems in early versions
  - Problems
    - Not reliable as completely subjective
    - Not valid as introspector is a non-typical user
    - Intuitions and introspection are often wrong
- Inspection methods help
  - Task centered walkthroughs
  - Heuristic evaluation

Cognitive Walkthrough

- Given:
  - a specification of the system (not neccesarily complete, but fairly detailed)
  - a description of the task the user is to perform on the system (representative for most users ...)
  - a complete, written list of the actions needed to complete the task
  - an indication of who the users are and what kind of experience and knowledge the evaluators can assume about them

Cognitive Walkthrough .

- Step through the action sequence and critique the system using questions:
  - Is the effect of the action the same as the user's goal at that point ?
  - Will users see that the action is available ?
  - Once users found the correct action, will they know it is the one they need ?
  - After the action is taken, will users understand the feedback they get ?

Conceptual Model Extraction

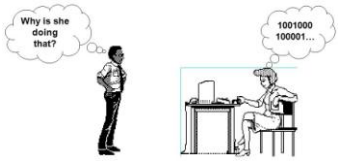
- How?
  - Show the user static images of
    - The prototype or screens during use
  - Ask the user explain
    - The function of each screen element
    - How they would perform a particular task
- What?
  - Initial conceptual model
    - How person perceives a screen the very first time it is viewed
  - Formative conceptual model
    - How person perceives a screen after its been used for a while
- Value?
  - Good for eliciting people's understanding before & after use
  - Poor for examining system exploration and learning

Direct Observations

- Evaluator observes users interacting with system
  - In lab:
    - User asked to complete a set of pre-determined tasks
  - In field:
    - User goes through normal duties
- Value
  - Excellent at identifying gross design/interface problems
  - Validity depends on how controlled/contrived the situation is

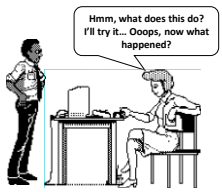
Simple Observation Method

- User is given the task
- Evaluator just watches the user
- Problem
  - Does not give insight into the user's decision process or attitude



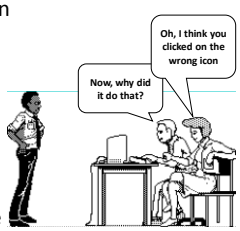
Think Aloud Method

- Users speak their thoughts while doing the task
  - What they are trying to do
  - Why they took an action
  - How they interpret what the system did
  - Gives insight into what the user is thinking
  - Most widely used evaluation method in industry
  - May alter the way users do the task
  - Unnatural (awkward and uncomfortable)
  - Hard to talk if they are concentrating



Constructive Interaction Method

- Two people work together on a task
  - Monitor their normal conversations
  - Removes awkwardness of think-aloud
- Co-discovery learning
  - Use semi-knowledgeable "coach" and novice
  - Only novice uses the interface
    - novice ask questions
    - coach responds
  - Gives insights into two user groups



Recording Observations

- How do we record user actions for later analysis?
  - Otherwise risk forgetting, missing, or misinterpreting events
  - Paper and pencil
    - Primitive but cheap
    - Observer records events, comments, and interpretations
    - Hard to get detail (writing is slow)
    - 2nd observer helps...
  - Audio recording
    - Good for recording think aloud talk
    - Hard to tie into on-screen user actions
  - Video recording
    - Can see and hear what a user is doing
    - One camera for screen, rear view mirror useful...
    - Initially intrusive



Coding Sheet Example

- Tracking a person's use of an editor

	General actions			Graph editing			Errors	
Time	text editing	scrolling	image editing	new node	delete node	modify node	correct error	miss error
09:00	X							
09:02				X				
09:05							X	
09:10					X			
09:13								

Interviews

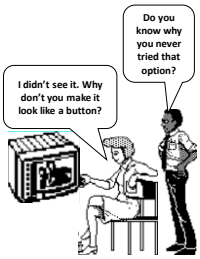
- Good for pursuing specific issues
  - Vary questions to suit the context
  - Probe more deeply on interesting issues as they arise
  - Good for exploratory studies via open-ended questioning
  - Often leads to specific constructive suggestions
- Problems:
  - Accounts are subjective
  - Time consuming
  - Evaluator can easily bias the interview
  - Prone to rationalization of events/thoughts by user
  - User's reconstruction may be wrong
  - Sometimes difficult to find people!

How to Interview

- Plan a set of central questions
  - A few good questions gets things started
    - Avoid leading questions
  - Focuses the interview
  - Could be based on results of user observations
- Let user responses lead follow-up questions
  - Follow interesting leads vs bulldozing through question list

Retrospective Testing Interviews

- Post-observation interview to
  - Perform an observational test
  - Create a video record of it
  - Have users view the video and comment on what they did
    - Clarify events that occurred during system use
    - Excellent for grounding a post-test interview
    - Avoids erroneous reconstruction
    - Users often offer concrete suggestions



Critical Incidence Interviews

- People talk about incidents that stood out
  - Usually discuss extremely annoying problems with passionate feeling
  - Not representative, but important to them
  - Often raises issues not seen in lab tests

Questionnaires and Surveys

- Questionnaires / Surveys
  - Preparation "expensive," but administration cheap
    - Can reach a wide subject group (e.g. mail)
  - Does not require presence of evaluator
  - Results can be quantified
- But
  - Only as good as the questions asked

Questionnaires and Surveys .

- How
  - Establish the purpose of the questionnaire
    - What information is sought?
    - How would you analyze the results?
    - What would you do with your analysis?
  - Do not ask questions whose answers you will not use!
  - Determine the audience you want to reach
  - Determine how would you will deliver / collect the questionnaire
  - On-line for computer users
  - Web site with forms
  - Surface mail
    - Pre-addressed reply envelope gives far better response

Styles of Questions

- Open-ended questions
  - Asks for unprompted opinions
  - Good for general subjective information
    - But difficult to analyze rigorously
- Can you suggest any improvements to the interfaces?

Styles of Questions .

- Closed questions
  - Restrict respondent's responses by supplying alternative answers
  - Makes questionnaires a chore for respondent to fill in
  - Can be easily analyzed
  - Watch out for hard to interpret responses!
    - Alternative answers should be very specific

Do you use computers at work:
O often O sometimes O rarely

vs

In your typical work day, do you use computers:
O over 4 hrs a day
O between 2 and 4 hrs daily
O between 1and 2 hrs daily
O less than 1 hr a day

Styles of Questions ..

- Scalar
  - Ask user to judge a specific statement on a numeric scale
  - Scale usually corresponds with agreement or disagreement with a statement

Characters on the computer screen are:

hard to read easy to read
1 2 3 4 5

Styles of Questions ...

- Multi-choice
  - Respondent offered a choice of explicit responses

How do you most often get help with the system? (tick one)
- on-line manual
- O paper manual
- O ask a colleague

Which types of software have you used? (tick all that apply)
- word processor
- O data base
- spreadsheet
- O compiler

Styles of Questions ....

- Ranked
  - Respondent places an ordering on items in a list
  - Useful to indicate a user's preferences
  - Forced choice

Rank the usefulness of these methods of issuing a command
(1 most useful, 2 next most useful..., 0 if not used
__2__ command line
__1__ menu selection
__3__ control key accelerator

Styles of Questions .....

- Combining open-ended and closed questions
  - Gets specific response, but allows room for user's opinion

It is easy to recover from mistakes:

disagree agree comment: undo facility is really helpful
1 2 3 4 5

Continuous Evaluation

- Monitor systems in actual use
  - Usually late stages of development
    - i.e. beta releases, delivered system
  - Fix problems in next release
- User feedback via gripe lines
  - Users can provide feedback to designers while using the system
    - Help desks
    - Bulletin boards
    - Email
    - Built-in gripe facility
  - Best combined with trouble-shooting facility
    - Users always get a response (solution?) to their gripes

Continuous Evaluation .

- Case/field studies
  - Careful study of "system usage" at the site
  - Good for seeing "real life" use
  - External observer monitors behavior
  - Site visits

Ethics

- Testing can be a distressing experience
  - Pressure to perform, errors inevitable
  - Feelings of inadequacy
  - Competition with other subjects
- Golden rule
  - Subjects should always be treated with respect

Ethics - Before the Test

- Don't waste the user's time
  - Use pilot tests to debug experiments, questionnaires etc
  - Have everything ready before the user shows up
- Make users feel comfortable
  - Emphasize that it is the system that is being tested, not the user
  - Acknowledge that the software may have problems
  - Let users know they can stop at any time
- Maintain privacy
  - Tell user that individual test results will be completely confidential
- Inform the user
  - Explain any monitoring that is being used
  - Answer all user's questions (but avoid bias)
- Only use volunteers
  - User must sign an informed consent form

Ethics - During the Test

- Don't waste the user's time
  - Never have the user perform unnecessary tasks
- Make users comfortable
  - Try to give user an early success experience
  - Keep a relaxed atmosphere in the room
  - Coffee, breaks, etc
  - Hand out test tasks one at a time
  - Never indicate displeasure with the user's performance
  - Avoid disruptions
  - Stop the test if it becomes too unpleasant
- Maintain privacy
  - Do not allow the user's management to observe the test

Ethics - After the Test

- Make the users feel comfortable
  - State that the user has helped you find areas of improvement
- Inform the user
  - Answer particular questions about the experiment that could have biased the results before
- Maintain privacy
  - Never report results in a way that individual users can be identified
  - Only show videotapes outside the research group with the user's permission

What you Now Know

- Debug designs by observing how people use them
  - Quickly exposes successes and problems
  - Specific methods reveal what a person is thinking
  - But naturalistic vs laboratory evaluations is a trade-off
- Methods:
  - Conceptual model extraction
  - Direct observation
    - Think-aloud
    - Constructive interaction
  - Query via interviews, retrospective testing and questionnaires
  - Continuous evaluation via user feedback and field studies
- Ethics are important

Questions

Acknowledgements

- Prof. Ing. Jiří Sochor