Week 3: Usability Evaluation

CS285 Usability Engineering
Evaluation - When?

• Ongoing with designing and building in users’ needs and goals – formative evaluation
• During prototyping - when changes can still be made
• End of design – summative evaluation (assess success or failure of finished product)
Evaluation - Why?

- To guide and inform design
- Determine requirements for design
- To capture usability problems / shortfalls
- To allow problems to be fixed
- To deal with the above early
- Help identify opportunities for new technology
What to evaluate

• Which features?
• What aspects?
• What user goals?

Usability goals

- have good utility
- easy to learn
- easy to remember how to use
- effective to use
- emotionally fulfilling
- entertaining
- rewarding
- supportive of creativity
- aesthetically pleasing
- motivating
- helpful
- fun
- satisfying
- efficient to use
- rewarding
- successful
- enjoyable
- fun

Usability Evaluation

• Evaluation Techniques (using a combination of)
  – Observations
  – Questionnaires
  – Interviews
  – User testing --> central concern

• To obtain objective performance data to show how usable a system/product is in terms of usability goals
Evaluation Approaches

• Empirical Approach
  – Involves working with users to gather data to analyse user performance in relation to proposed system
  – Observations, questionnaires, interviews, and experiments (testing)

• Analytical Approach
  – Predict performance based on a model
  – e.g. analysis of cash dispenser based on number of key strokes required, time needed to press a key, time needed to react, etc.

• Combination of both
Evaluation Paradigms

- “Quick and dirty” evaluation
- Usability testing
- Field studies
- Predictive evaluation
Evaluation Techniques

- Observing users
- Asking users
- Asking experts
- User Testing
- Modelling user’s task performance
“Quick and dirty” evaluation

• Usually descriptive data
• Informal
• Done in a short space of time
• Feedback informs design
• Can be done at any stage
• Emphasis on fast input
• Real users are involved – empirical
• No users – analytical (expert heuristic)
Usability Testing

- A systematic approach to evaluate user performance to inform and improve usability design
- Measure performance of typical users in controlled laboratory-like conditions
- Should be carefully planned and executed within limits of real world constraints
- Less rigorous than research experiments (an applied form of experimentation)
- Typically 5 – 12 users, if less --> “quick and dirty”
Usability Testing

• *Capture*: collecting usability data, such as task completion time, errors, guideline violations, and subjective ratings
  – Observations (including video)
  – Logging interactions (software)
  – Satisfaction questionnaires/interviews

• *Analysis*: interpreting usability data to identify usability problems in the interface;

• *Critique*: suggesting solutions or improvements to mitigate problems.
Field Studies

• Done in natural settings
• Field Studies – Outsider
  – Explicit observing & recording as it happens
  – As outsider looking on
• Field Studies – Insider (possibly participant)
  – To explore the details of what happens in a particular social setting --> Ethnography
• Data analyzed qualitatively and / or quantitatively
Predictive Evaluation

• Heuristic evaluation
  – experts review the software design/product guided by tried and tested heuristics, applying their knowledge of typical users (& tasks)
  – Process is quick, relatively inexpensive
  – Has limitations, care needed

• GOMS models (goals, operators, methods and selection rules)

• Users need not be present or tested
Alternative Techniques

• Via networking, Internet, & video conferencing
  – Remote-control evaluation
  – Instrumented remote evaluation (also real-time)
  – Semi-instrumented remote evaluation (users trained to identify and “tag” usability problems during their normal usage of an application)

• Drawbacks
• Benefits
DECIDE Evaluation Framework

- Determine the overall goals
- Explore the specific questions
- Choose the paradigm and technique
- Identify practical issues
- Deal with ethical issues
- Evaluate, interpret, analyze, and present the data
Determine the goals

- Goals guide an evaluation
  - Influence the choice of approach, paradigm, & technique
- Who wants the evaluation and why
- Examples of high level goals statements
  - Identify the metaphor on which to base design
  - Check to ensure final interface is consistent
  - Investigate degree to which a technology affects current working practices
Explore the questions

• Identify questions to fulfil/satisfy goals
• To make the goals operational
• To make the evaluation specific, use sub-questions, sub–sub questions, & so on
• Example: Think of questions that will enable you to find out why customers prefer to purchase paper airline tickets rather than e-tickets
Choose paradigm & technique

• Paradigm determines techniques
• Combination of techniques to obtain different perspectives
• Consideration of equipment and/or expertise available
• Consideration of time and expense (practical issues)
Identify Practical Issues

• Users
  – Appropriate users (screened), representative sample
  – How they will be involved?
• Facilities and equipment
  – The effects of using an equipment (camera)
• Schedule and budget constraints
• Available expertise
Deal with ethical issues

- People’s privacy should be protected
- Obtain permissions (informed consent forms)
- Respect confidentiality (anonymity) of all data collected
- Keep users informed of evaluation purposes and what to expect
- Option to withdraw or terminate
Evaluate Interpret Present Data

• Qualitative analysis – interpreted to tell a story (with incidents, anecdotal illustration)
• Qualitative analysis – categorized to identify incidents or patterns (content analysis, discourse analysis)
• Quantitative analysis – statistical treatment of interaction and video log (including categorized data)
How to observe users

• Direct observation
  – In controlled environments --> think-aloud technique
  – In the field --> prepare a framework (checklist) for organizing and for data collection

• Participant Observation

• Indirect Observation
  – Diaries
  – Interaction Logging
Asking users: interviews

- Unstructured --> open-ended, conversations focusing on a particular topic, may be in-depth
- Structured --> poses predetermined questions, short, clearly worded, closed (precise answer)
- Semi-structured --> both closed and open questions, use prompts and probes, a basic script to ensure consistency
- Types: face-to-face, group, telephone, online
Asking users: questionnaires

- Make questions clear & specific
- Ask closed questions & offer a range of answers
- Include a “no-opinion” option if seeking opinions
- Order questions logically
- Do not use scale ranges that overlap
- Avoid jargon, consider different versions for different target groups
- Provide clear instructions on how to fill it out
- Keep it compact with a adequate white space
Asking experts: inspections

• Heuristic evaluation (Nielsen, 2001)
  – Experts use a set of usability principles, may need to tailor one to suit product/system
  – Experts role play typical users
  – Experts record problems encountered
  – 5 experts usually identify 75% of problems
  – *Expert crit* - critiquing interactive devices
  – Discount evaluation → quick and inexpensive
  – Never a replacement for user testing
Ten Usability Heuristics

- Nielsen, J. (www.useit.com/papers/heuristic)
  - Visibility of system status
  - Match between system and the real world
  - User control and freedom
  - Consistency and standards
  - Help users recognise, diagnose, and recover from errors
  - Recognition rather than recall
  - Flexibility and efficiency of use
  - Aesthetic and minimalist design
  - Help and documentation
  - Error prevention
Asking experts: walkthroughs

• Cognitive walkthroughs
  – Evaluator walks through the action sequences for each task in the context of a scenario
    • Will the correct action be sufficiently evident to the user?
    • Will the user notice the correct action is available?
    • Will the user interpret the response from the action correctly?
  – Focuses on user’s problems in detail, laborious and time consuming
Asking experts: walkthroughs

- Pluralistic walkthroughs
  - Panel of evaluators; users, designers, usability experts, etc.
  - Record individually the actions to walkthrough action sequences of a scenario
  - Panel discuss, compare, comment, review, recommend

- Participatory approach → multidisciplinary
- Not easy getting everybody together
Experiments

• To answer a question or test a hypothesis that predicts a relationship between two or more events known as variables

• Designs
  – 1x2 design; type font (variable) with 2 conditions
  – 2x2 design; versions (2 variables) with levels of expertise (2 conditions)

<table>
<thead>
<tr>
<th>Font</th>
<th>Times</th>
<th>Helvetica</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginners</td>
<td>Beginners</td>
<td>Experts</td>
</tr>
<tr>
<td>Experts</td>
<td>Experts</td>
<td></td>
</tr>
</tbody>
</table>
## Experiments: Participant

<table>
<thead>
<tr>
<th>Design</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different participants</td>
<td>No order effects</td>
<td>Many participants needed. Individual differences among participants are a problem. Can be offset to some extent by randomly assigning to groups</td>
</tr>
<tr>
<td>Same participants</td>
<td>Eliminates individual differences between experimental conditions</td>
<td>Need to counterbalance to avoid ordering effects</td>
</tr>
<tr>
<td>Matched participants</td>
<td>Same as different participants but the effects of individual differences are reduced</td>
<td>Can never be sure that subjects are matched across variables.</td>
</tr>
</tbody>
</table>
Modelling users

• GOMS model (goals, operators, methods, and selection rules)
  – Goals: a particular state the user wants to achieve
  – Operators: cognitive processes and physical actions needed to attain goal
  – Methods: learned procedures (exact sequences of steps) to attain goal
  – Selection rules: if more than one method, rules route to a method
Modelling users

• Keystroke model
  – Estimate/Obtain time to execute operators
  – Predict time to execute a given task by summing the approx time of each action:
    \[ T_{\text{execute}} = T_K + T_P + T_H + T_D + T_M + T_R \]

• Fitts’ Law (1954)
  – Predicts the time it takes to reach a target with a pointing devise
    \[ T = k\log_2(D/S + 0.5), \quad k \sim 100\text{msec} \]
    \[ T = \text{time to move hand to a target, } D = \text{distance between hand and target, } S = \text{size of target} \]
Modelling Users

• Benefits and limitations
  – Make predictions about predictive behaviour only
  – People are unpredictable in the way they behave
  – Difficult to evaluate how people will use systems in real-world contexts
  – Provide estimates for comparing the efficiency of different methods of completing tasks.