Questions?

• Project #1
  – You need to help users figuring what they want
    • *Example: keyboard in the living room*
    • *But avoid Stockholm syndrome*
  – You need to prioritize their request
    • *Too many features might be confusing*
    • *You cannot always serve all masters*
  – You need to protect users anonymity
    • *Use of persona will help*
  – Grading requests
    • *Please be courteous*

• Project #2
  – This Thursday
User evaluation

- The design cycle

- Different style of user testing
  - Matching the need of different design phase
  - Qualitative or Quantitative
Evaluation criteria

- Learning time and Retention
- Speed
- Error rate
- Subjective satisfaction
- …

- Tradeoffs are often needed
Qualitative approach

• Gather users perception of the interaction

• Methods
  – Interviews, questionnaires and surveys
  – Introspection
    • *Walkthroughs*
  – Direct observation
    • *Direct observation*
    • *Thinking aloud*
    • *Constructive interaction*
Qualitative approaches outcome

• High level effects
  – Task flow problems
  – Task description problems
  – Contextual findings
    • Conflict with social pattern, ...
    • Two hands needed but only one available

• Pros and Cons
  – Apply to a real situation
    • Good external validity
  – Difficult to generalize
    • Poor control of independent variables
  – Often subjective data
Quantitative approach

- Gather (performance) measurements

Methods
  - User events collection
    - Mouse clicks, keys pressed, ...
    - Data collected during system use
      - Google, Amazon
  - Controlled experiments
    - Set forth a testable hypothesis
    - Manipulate one or more independent variable
    - Observe effect on one or more dependent variable
    - Can be reproduced by others
Quantitative approach outcome

• Low level effects
  – Patterns of use
  – Menu selection method A faster than method B

• Pros and cons
  – Objective measurements
    • *Good internal validity*
  – Real world implications sometime difficult to foresee
  – Effects might be dwarfed in real world settings
    • *3.05s versus 3.00s?*
Questioning measurements

• Are they reliable?
  – Will repeating the experiment deliver the same result?
  – Does the experiment take into account variations between subjects?
    • Need for testing a sample of subjects

• Are they valid?
  – Does the experiment reflects target use?
    • Were users typical?
    • Were tasks typical?
    • Was the setting realistic?
    • Was the experience biased?
Are results significant?

• Statistical significance
  – Comparing to the null hypothesis: “There is no effect”
  – Type I errors are the most disruptive

<table>
<thead>
<tr>
<th>Researcher’s Decision</th>
<th>Actual Situation: Null Hypothesis is</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
</tr>
<tr>
<td>Accept the null hypothesis</td>
<td>Correct decision</td>
</tr>
<tr>
<td>Reject the null hypothesis</td>
<td><strong>Type I error</strong></td>
</tr>
</tbody>
</table>

• Design significance?
  – 3.00s versus 3.05s?
Treating subjects with respect

• Testing is a distressing experience
  – Pressure to perform
  – Feeling of inadequacy
  – Looking like a fool in front of your peers, your boss,…

• Follow human subject protocols
  – Individual test results will be kept confidential
  – Users can stop the test at any time
  – Users are aware (and understand) the monitoring technique
  – Their performance will have not direct implication on their life
  – Records will be made anonymous
    • Videos

• Use standard informed consent form
  – Especially for quantitative tests
  – Be aware of legal requirements
Conducting the experiment

• Before the experiment
  – Have them read and sign the consent form
  – Explain the goal of the experiment
    • In a way accessible to users
    • Be careful about the demand characteristic
    • Answer questions

• During the experiment
  – Stay neutral
    • Never indicate displeasure with users’ performance

• After the experiment
  – Debrief users
    • Inform users about the goal of the experiment
  – Answer any questions they have
Managing subjects

• Don’t waste users time
  – Use pilot tests to debug experiments, questionnaires, etc…
  – Have everything ready before users show up

• Make users comfortable
  – Keep a relaxed atmosphere
  – Allow for breaks
  – Pace tasks correctly
  – Stop the test if it becomes too unpleasant