USABILITY SPECIFICATIONS

Chapter 8

TOPICS:

• What they are
• Usability specification table
• Benchmark task descriptions
• User errors in usability specifications
• Usability specifications and managing the UE process
• Team exercise on usability specifications

Copyright © 2001 H. Rex Hartson and Deborah Hix.

All rights reserved. No part of this material may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without prior written permission of its authors.
INTRODUCTION TO USABILITY SPECIFICATIONS

• Revisiting the usability engineering life cycle
USABILITY SPECIFICATIONS

Another piece of the puzzle...

• Introducing Envision
  
* Videoclips to be used as examples of process activities

* Envision: a digital library of computer science literature

* Search results are presented in graphical scatterplot

* User-controllable visualization along axes of scatterplot

* Videoclip: Envision prototype
USABILITY SPECIFICATIONS

Usability is not something warm, fuzzy, but is quantitative

- Quantitative usability goals against which user interaction design is measured

- Target levels for usability attributes
  
  * Operationally defined metrics for a usable interaction design

  If you can't measure it, you can't manage it

  * Management control for usability engineering life cycle

  * Indication that development process is converging toward a successful interaction design

  * Establish as early in process as feasible
USABILITY SPECIFICATIONS

• Tie usability specifications to early usability goals

  * E.g., for early goal of walk-up-and-use usability, base usability specification on initial task performance time

• All project members should agree on usability specification attributes and values

  Get customers and users involved in setting levels
USABILITY SPECIFICATION DATA

• Usability specifications based on

* Objective, observable user performance

* Subjective user opinion and satisfaction

- Subjective preferences may reflect users desire to return to your Web site, but flash and trash soon bores and irritates

• Objective and subjective usability specifications can both be quantitative
**USABILITY SPECIFICATION TABLE**

Credit to: [Whiteside, Bennett, & Holtzblatt, 1988]

<table>
<thead>
<tr>
<th>Usability Attribute</th>
<th>Measuring Instrument</th>
<th>Value to be Measured</th>
<th>Current Level</th>
<th>Target Level</th>
<th>Observed Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Usability attribute* — what general usability characteristic is to be measured

For Y2K Calender: initial performance, since want good 'walk-up-and-usability' w/o training or manuals

- May need separate usability attributes for each user class

- Some quantitative usability attributes

*Objective

- Initial performance

- Longitudinal (experienced, steady state) performance

- Learnability

- Retainability
USABILITY SPECIFICATION TABLE

• Some other quantitative usability attributes

* Subjective
  - Initial impression
  - Longitudinal satisfaction
USABILITY SPECIFICATION TABLE

<table>
<thead>
<tr>
<th>Usability Attribute</th>
<th>Measuring Instrument</th>
<th>Value to be Measured</th>
<th>Current Level</th>
<th>Target Level</th>
<th>Observed Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Measuring instrument — vehicle by which values are measured for usability attribute, the thing that generates the data (e.g., task generates timing data, questionnaire generates preference data)

Benchmark 1: Schedule a meeting with Dr. Ehrich for four weeks from today at 10AM in 133 McBryde, about the HCI research project.

• What tasks should be included?

* Representative, frequently performed tasks

* Common tasks – 20% that account for 80% of usage

* Critical business tasks – not frequent, but if you get it wrong, heads can roll
**USABILITY SPECIFICATION DATA**

- User performance for specific benchmark tasks
  
  objective

- Questionnaire score
  
  subjective (opinion)

- Example question from QUIS:

<table>
<thead>
<tr>
<th></th>
<th>terrible</th>
<th>wonderful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall reaction to user interface</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>frustrating</th>
<th>satisfying</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 1 2 3 4 5 6 7 8 9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>uninteresting</th>
<th>interesting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 1 2 3 4 5 6 7 8 9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>dull</th>
<th>stimulating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 1 2 3 4 5 6 7 8 9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>difficult</th>
<th>easy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 1 2 3 4 5 6 7 8 9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>
BENCHMARK TASK DESCRIPTIONS

• Developing benchmark task descriptions
  * Representative, frequently performed tasks
  * Clear, precise, repeatable instructions
  * *What* task to do, not *how* to do it
  * Clear start and end points for timing

Not good: Display last week's appts.

E.g., "access a certain kind of info" better than "submit query", better match to work activity

* Adapt scenarios already developed for design
  - Clearly an important task to evaluate
  - Remove information about how to do it
BENCHMARK TASK DESCRIPTIONS

• Developing benchmark task descriptions
  (continued)

  * Start with fairly simple tasks, then progressively increase difficulty

    E.g., add an appt, then add an appt 60 days from now, or add multiple recurring appts.
    E.g., move an appt from one month to another

  * Avoid large amounts of typing if typing skill is not being evaluated

  * To evaluate error recovery, benchmark task can begin in error state
BENCHMARK TASK DESCRIPTIONS

• Developing benchmark task descriptions (continued)

* Tasks should include navigation

  E.g., 'add appt one month from today' causes crossing month boundary independent of absolute date, BUT...

* Task wording should be unambiguous

  Above ex. is confusing; also, wording of tasks can influence task performance (4 wks -- users move by week; 1 mo -- users move by month)

* Don’t use words in benchmark tasks that appear specifically in interaction design

  E.g., “Find first appt...”, when there’s a button labeled “find”. Instead, word task like: “search for” or “locate”

* Typically put each benchmark on a separate sheet of paper

* Typical number of benchmark tasks: Enough for reasonable, representative coverage

* Example for Y2K Calendar: Add an appointment with Dr. Kevorkian for 4 weeks from today at 10AM concerning your flu shot.
# USABILITY SPECIFICATION TABLE

<table>
<thead>
<tr>
<th>Usability Attribute</th>
<th>Measuring Instrument</th>
<th>Value to be Measured</th>
<th>Current Level</th>
<th>Target Level</th>
<th>Observed Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial performance</td>
<td>&quot;add appt&quot; task per Benchmark 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Value to be measured* — metric for which usability data values are collected

- Time to complete task
- Number of errors
- Frequency of help and documentation use
- Time spent in errors and recovery
- Number of repetitions of failed commands
- Number of times user expresses frustration or satisfaction
- Number of commands, mouse clicks, or other user actions to perform task(s)
## USABILITY SPECIFICATION TABLE

<table>
<thead>
<tr>
<th>Usability Attribute</th>
<th>Measuring Instrument</th>
<th>Value to be Measured</th>
<th>Current Level</th>
<th>Target Level</th>
<th>Observed Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial performance</td>
<td>&quot;add appt&quot; task per Benchmark 1</td>
<td>Length of time to add appt on first trial</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* **Current level** — present value of usability attribute to be measured

- Level of performance for current version of system for measuring instrument (when available)
- Baseline to help set target level, from:
  * Manual system (performing task without computer)
  * Automated system (existing or previous version)
  * Competitor system
  * Developer performance (for expert, longitudinal use)
## USABILITY SPECIFICATION TABLE

<table>
<thead>
<tr>
<th>Usability Attribute</th>
<th>Measuring Instrument</th>
<th>Value to be Measured</th>
<th>Current Level</th>
<th>Target Level</th>
<th>Observed Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial performance</td>
<td>&quot;add appt&quot; task per Benchmark 1</td>
<td>Length of time to add appt on first trial</td>
<td>20 secs. (competitor system)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Target level — value indicating unquestioned usability success for present version

- Minimum acceptable level of user performance
- Determining target level values

* Usually acceptable improvement over current level
USABILITY SPECIFICATION TABLE

- *Observed results* — actual values from evaluation with users, filled in after evaluation sessions

- More example usability specifications

<table>
<thead>
<tr>
<th>Usability Attribute</th>
<th>Measuring Instrument</th>
<th>Value to be Measured</th>
<th>Current Level</th>
<th>Target Level</th>
<th>Observed Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial performance</td>
<td>&quot;add appt&quot; task per Benchmark 1</td>
<td>Length of time to add appt on first trial</td>
<td>20 secs. (competitor system)</td>
<td>15 secs.</td>
<td></td>
</tr>
</tbody>
</table>
USER ERRORS IN USABILITY SPECIFICATIONS

• Determining errors

* Deviation from any correct path to accomplish task (except, for example, going to Help)

* Only situations that imply usability problems

* Do not count "oops" errors, doing wrong thing when knew it was wrong

E.g., using wrong month on calendar totally by accident
USER ERRORS IN USABILITY SPECIFICATIONS

• Examples of errors

* Selecting wrong menu, button, icon, etc. when user thought it was the right one

* Double clicking when single click is needed, and vice versa

* Using wrong accelerator key

* Operating on the wrong interaction object (when user thought it was the right one)
  
  E.g., working on wrong month of calendar because they couldn't readily see month's name

* Usually not typing errors
CREATING USABILITY SPECIFICATIONS

• Usability evaluation design driven by usability goals

* First determine usability goals

- In terms of user class, task context, special tasks, marketing needs

  Example: Increase number of times user can perform task X, based on current version, implies:

  - Example: Reduce amount of time for novice user to perform task X in Version 2.0

  - Be as specific as possible

* Then quantify as usability specification(s)

  - Example: Currently 35 seconds to perform task X ("current level"); reduce to 25 seconds ("target level")
CREATING USABILITY SPECIFICATIONS

• What are constraints in user or work context? How can setting be more realistic? (Ecological validity)

* Usability lab can be "sterile work environment"

* Does task require telephone, more than one person or role?

* Does task involve background noise?

* Does task involve interference, interruption?

  Phone can be used for this, too.

• Experimental design must take into account trade-offs among usability characteristics, user groups

* Watch out for potential trade-off between learnability for new users and performance power for experienced users
**USABILITY SPECIFICATION TABLE**

- Some additional example usability specifications

<table>
<thead>
<tr>
<th>Usability Attribute</th>
<th>Measuring Instrument</th>
<th>Value to be Measured</th>
<th>Current Level</th>
<th>Target Level</th>
<th>Observed Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial performance</td>
<td>&quot;add appt&quot; task per Benchmark 1</td>
<td>Length of time to add appt on first trial</td>
<td>20 secs. (competitor system)</td>
<td>15 secs.</td>
<td></td>
</tr>
<tr>
<td>Initial performance</td>
<td>&quot;add appt&quot; task per Benchmark 1</td>
<td>Number of errors</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
USABILITY SPECIFICATIONS HELP MANAGE THE USABILITY ENGINEERING PROCESS

• This is the *control* of usability engineering life cycle

* Quantifiable end to process

* Accountability

* Stop iterating when target level usability specifications are met

* It’s expected that you will not meet all usability target levels on first iteration

- If usability target levels are met on first iteration, they may have been too lenient

- Point is to uncover usability problems
USABILITY SPECIFICATIONS HELP MANAGE THE USABILITY ENGINEERING PROCESS

• Bottom line: **Good engineering judgment is important**

* For setting levels (especially "target" level)
* For knowing if specifications are "reasonable"

• **Videoclip**: Setting usability specifications
TEAM EXERCISE: USABILITY SPECIFICATIONS

• **Goal:**

* To gain experience in writing precise, measurable usability specifications

• **Activities:**

* Produce three usability specifications, two based on objective measures, and one based on subjective measures.

* For each specification for objective measures, write a brief but specific benchmark task description for participants to perform. Write two different benchmark task descriptions, each on a separate sheet. Have them be a little complicated, including some navigation.
TEAM EXERCISE: USABILITY SPECIFICATIONS

* Specifications with objective measures should be evaluable, via the benchmark tasks, in a later class exercise, on formative evaluation. Develop tasks that you can "implement" in your next exercise, to build a rapid prototype.

* The specification for subjective measure should be based on the questionnaire supplied. Select 3 or 4 items from questionnaire.

• Cautions and hints:

* Don’t spend any time on design in this exercise; there will be time for detailed design in the next exercise.

* Don't plan to give users any training.
TEAM EXERCISE: USABILITY SPECIFICATIONS

»» Three usability specifications, in the form on a transparency

»» Questionnaire question numbers included in subjective specification

»» Benchmark task descriptions, each on a separate sheet of paper

• *Completed by:*

  About 30 minutes; 40 mins. max.