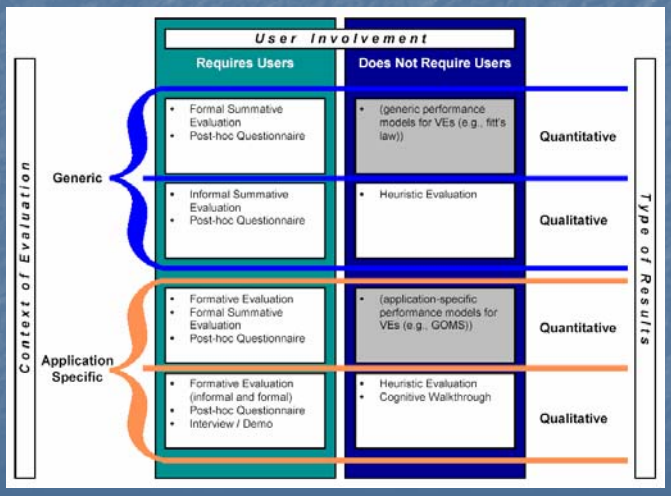
3D User Interface Evaluation II

Lecture #15: Evaluating 3DUIs – Part II

Spring 2008

Joseph J. LaViola Jr.

Usability Evaluation in 3DUIs



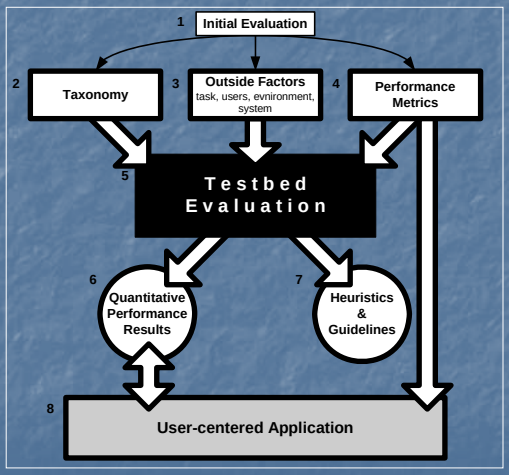
Classification Shortcoming

- Does not tell you “when” a method should be applied
- Does not tell you “how” to apply more than one method
- 3DUI evaluation models
  - Testbed evaluation
  - Sequential evaluation

Testbed Evaluation Framework

- Developed by Bowman and Hodges (1999)
- Empirically evaluate techniques outside of applications
- Components
  - initial evaluation
  - taxonomy
  - outside factors
  - performance metrics
  - testbed evaluation
  - application and generalization of results

Testbed Evaluation

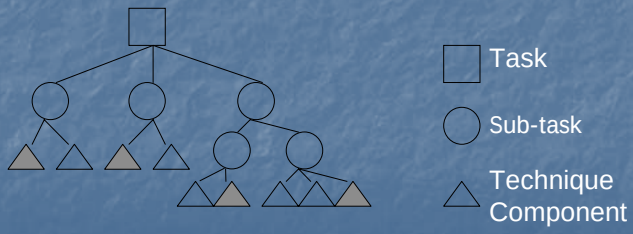


Testbed Evaluation – Initial Evaluation

- Gain intuitive understanding of generic interaction tasks and current technologies
- Experience and user observation
- Used for
  - building taxonomy
  - identifying outside factors
  - finding performance metrics

Testbed Evaluation – Taxonomy

- Develop taxomony of interaction techniques for interaction task in question
- Can use task-subtask approach



Testbed Evaluation – Outside Factors

- Cannot evaluate in a vacuum
- Need to take other factors into account
- Categories
  - task characteristics
  - environment characteristics
  - user characteristics
  - system characteristics

Testbed Evaluation – Metrics

- Objective measures
  - speed
  - accuracy
- Subjective measures
  - ease of use
  - ease of learning
  - frustration
  - etc...

Testbed Evaluation – The Testbed

- Allows generic, generalizable , and reusable evaluation
- Testbed
  - examines all aspects of a task
  - evaluates each technique component
  - considers outside influences
  - has good metrics
- Normally use formal, factorial experimental designs

Testbed Evaluation – Results

- Produces set of results or models that characterize an interaction technique for a given task
- Usability in terms of multiple performance metrics
- Results become part of a performance database for task
- Results can be generalized into heuristics or guidelines
- Apply to 3D applications

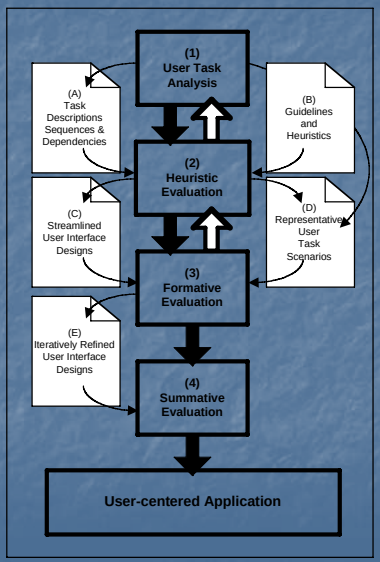
Testbed Evaluation Experiments

- Travel testbed (Bowman, Davis, et al. 1999)
  - compared seven different travel techniques
  - naïve and primed search
  - 44 subjects tested
- Selection/Manipulation testbed (Bowman and Hodges 1999)
  - compared nine different interaction techniques
  - 48 subjects
- Produced unexpected and intersting results (see papers for details)

Sequential Evaluation

- Developed by Gabbard, Hix, and Swan (1999)
- Usability engineering approach
- Evolved from existing GUI/2D evaluation methods
- Addresses both design and evaluation
- Employs
  - application specific guidelines
  - domain specific representative users
  - application specific user tasks

Sequential Evaluation



Sequential Evaluation – Example

- Applied to Dragon system
- Several evaluations performed in 9-month period
  - one to three users
  - two to three evaluators
  - Four cycles
- Guideline-based evaluation
- Summative evaluation
  - major study
  - four factors (2 x 2 x 3 x 2)
- See
  - Hix et al. (1999)
  - Hix and Gabbard (2002)

Comparison of Approaches

- Goals
  - Testbed – finding generic performance characteristics
  - Sequential – better UI for particular application
- Costs
  - Testbed – difficult experimental design, large numbers of trials and subjects
  - Sequential – multiple evaluators, significant time investment

3D Usability Evaluation

Things To Consider

Formality of Evaluation

- Formal: independent & dependent variables, statistical analysis, strict adherence to procedure, hold constant all other variables, usually done to compare multiple techniques or at the end of the design process
- Informal: looser procedure, often more qualitative, subject comments very important, looking for broad usability issues, usually done during the design process to inform redesign

What is Being Evaluated?

- Application:
  - Prototype - consider fidelity, scope, form
  - Complete working system
  - Controlled experiments are rare
- Interaction techniques / UI metaphors
  - Can still evaluate a prototype
  - More generic context of use
  - Formal experiments more often used
- Consider “Wizard of Oz” evaluation

Subjects / Participants

- How many?
- What backgrounds?
  - technical vs. non-technical
  - expert vs. novice VE users
  - domain experts vs. general population
- What age range?
- Recruiting
  - flyers
  - email/listservs/newsgroups
  - psychology dept.
  - CS classes

Number of Evaluators

- Multiple evaluators often needed for 3DUI evaluations
- Roles
  - cable wrangler
  - software controller
  - note taker
  - timer
  - behavior observer
  - ...

Procedure

- Welcome
- Informed consent
- Demographic/background questionnaire
- Pre-testing
- Familiarize with equipment
- Exploration time with interface
- Tasks
- Questionnaires / post-testing
- Interviews

- Subject "packets" are often useful for organizing information and data
- Pilot testing should be used in most cases to:
  - "debug" your procedure
  - identify variables that can be dropped from the experiment

Instructions

- How much to tell the subject about purposes of experiment?
- How much to tell the subject about how to use the interface?
- Always tell the subject what they should try to optimize in their behavior.
- If using think-aloud protocol, you will have to remind them many times.
- If using trackers, you will have to help users “learn” to move their heads, feet, and bodies – it doesn’t come naturally to many people.
- Remind subjects you are NOT testing them, but the interface.

Formal Experiment Issues

- Choosing independent variables
- Choosing dependent variables
- Controlling (holding constant) other variables
- Within- vs. between-subjects design
- Counterbalancing order of conditions
- Full factorial or partial designs

Independent Variables

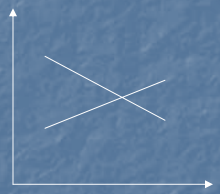
- Main variable of interest (e.g. interaction technique)
- Secondary variables
  - task characteristics
  - environment characteristics
  - system characteristics
  - user characteristics

Metrics (dependent variables)

- Task performance time
- Task errors
- User comfort (subjective ratings)
- Observations of behavior (e.g. strategies)
- Spoken subject comments (e.g. preferences)
- Surveys/questionnaires
- Interviews

Data Analysis

- Averages (means) of quantitative metrics
- Counts of errors, behaviors
- Correlate data to demographics
- Analysis of variance (ANOVA)
- Post Hoc analysis (t-tests)
- Visual analysis of trends (esp. learning)



- *Interactions between variables* are often important
- Expect high variance in 3DUI interaction studies

Analysis Tools

- SPSS, SAS, etc.
  - full statistical analysis packages
  - parametric and non-parametric tests
  - test correction mechanisms (e.g., Bonferroni)
- Excel
  - basic aggregation of data
  - Correlations
  - confidence intervals
  - graphs
- Matlab, Mathematica

Next Class

- 3DUI evaluation example
- Readings
  - 3DUI Book – Chapter 11, 367-384