CSE 190: 3D User Interaction

Lecture #16: 3D UI Evaluation

Jürgen P. Schulze, Ph.D.

Announcements

- Final Exam
  - Tuesday, March 19th, 11:30am-2:30pm
  - Closed book
  - See new section on course web page.
- CAPE
  - Web site closes March 18 at 8am
  - Responses to all surveys are completely anonymous.
  - Only a summary of results is provided to the academic department and the course instructor.
  - This summary is provided only after final grades are posted.
  - A minimum number of three evaluations must be submitted by students for summaries to made available.
- Please return borrowed webcams, Hydras, Kinects

Paper Presentations Next Lecture

- Joey: TBD

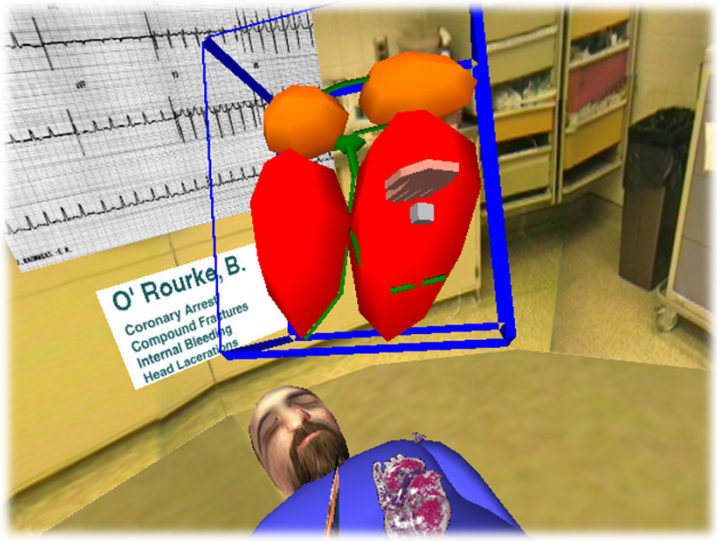
Paper Presentations Today

- Bryan: Impossible Spaces: Maximizing Natural Walking in Virtual Environments with Self-Overlapping Architecture
- Arick: Augmented perception of satiety: controlling food consumption by changing apparent size of food with augmented reality

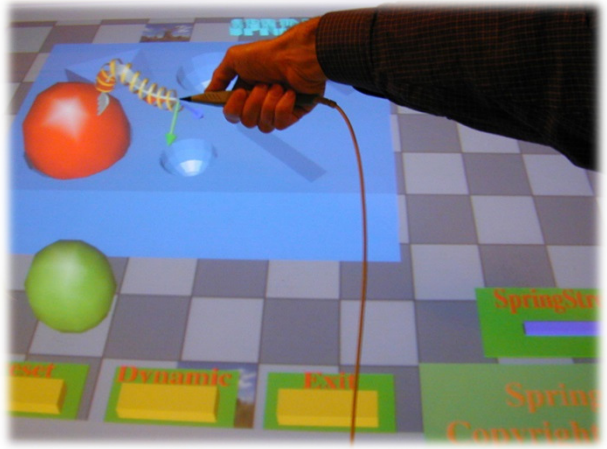
More on 3D UI Design Strategies

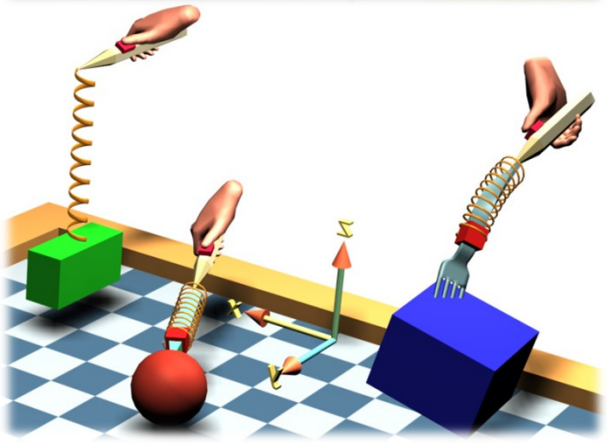
Designing for Humans – Feedback Substitution

- Cannot always support all sensory feedback dimensions
- Typical approach is to substitute



Highlighting object about to be selected

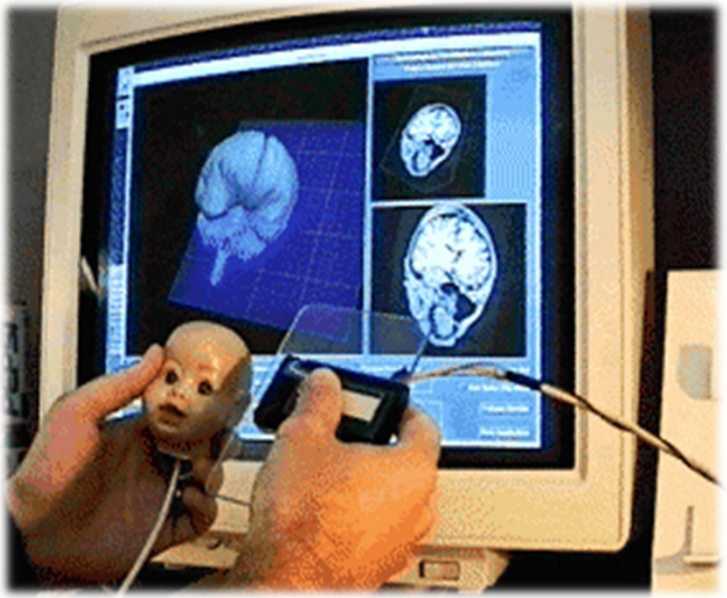




Spring Manipulation Tools, Michal Koutek, TU Delft

Designing for Humans – Passive Haptics

- Match shape and appearance of virtual object with physical prop
  - User both sees and feels
- Advantages
  - Inexpensive haptic/tactile feedback
  - Establish perceptual frame of reference
- Disadvantages
  - Scalability
  - Performance improvements have not yet been measured



Designing for Humans – Constraints

- Constraints:
  - Are a relation between variables that must be satisfied
    - Example: a line should stay horizontal
  - Define geometrical coherence of scene
  - Can make interaction simpler and improve accuracy

Designing for Humans – Constraint Types

- Physically realistic constraints
  - Collision detection and avoidance
  - Gravity
  - Application dependent
- DOF reduction
  - Simplify interaction (example: constrain travel to ground)
- Dynamic alignment tools
  - Grids and snapping, guiding surfaces
- Intelligent constraints
  - Deal with semantics
    - Example: lamp can only stand on horizontal surfaces

Designing for Humans – Two Handed Control

- Also known as bimanual input
- Transfer everyday manipulation experiences to 3DUI
- Can increase user performance on certain tasks
- Active topic of research

Designing for Humans – Guiard's Framework

- Tasks are
  - Unimanual: throwing darts
  - Bimanual symmetric
    - Synchronous: pulling a rope
    - Asynchronous: typing on keyboard
  - Bimanual asymmetric (cooperative): holding a cell phone with one hand, operating it with the other
- Division of labor (hand roles) for asymmetric scenario:
  - Nondominant hand dynamically adjusts spatial frame of reference for dominant hand
  - Dominant hand produces precision movements, nondominant hand performs gross manipulation
  - Manipulation is initiated by nondominant hand

Designing for Different User Groups

- Age
- Prior 3DUI experience
- Physical characteristics: arm length, etc.
- Perceptual, cognitive, motor capabilities
  - Color recognition
  - Stereo vision
  - Spatial abilities

Designing for User Comfort

- Weight of equipment
- Keep users in proper physical space
- Hygiene and public installations
- Keep sessions short (30-45min max) to prevent sickness, fatigue

3DUI Design

- Two main strategies
  - Designing for humans
    - Match design to human strengths
  - Inventing 3D interaction techniques
    - Creative exploration of 3D UIs

Inventing 3D User Interfaces

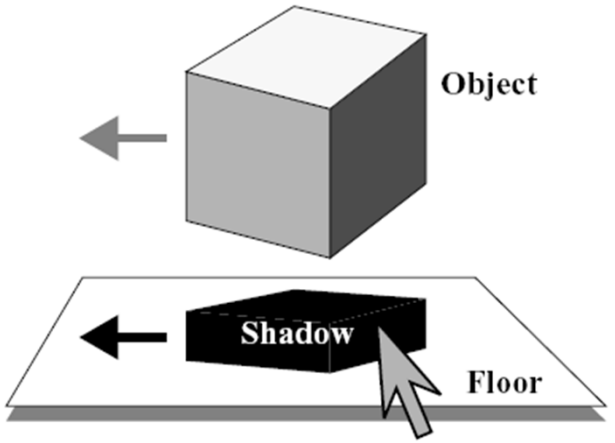
- Realism (or isomorphism)
  - Borrowing from real world
- Magic (or non-isomorphism)
  - Deviating from the real world and introducing artificial, magic techniques
- Continuum between realism and magic

Inventing 3DUIs – Simulating Reality

- Tried and true approach
  - replicate world as close as possible
  - bring in certain elements
- Important for simulation applications
  - flight simulators
  - medical training
  - phobia treatment
- Dependent on application
- Advantages
  - User already knows how to do it from everyday experience
  - Can be implemented on the basis of designer intuition
- Disadvantages
  - Limitations of technology do not allow exact realism
  - Introduces limitations of the physical world into the virtual world

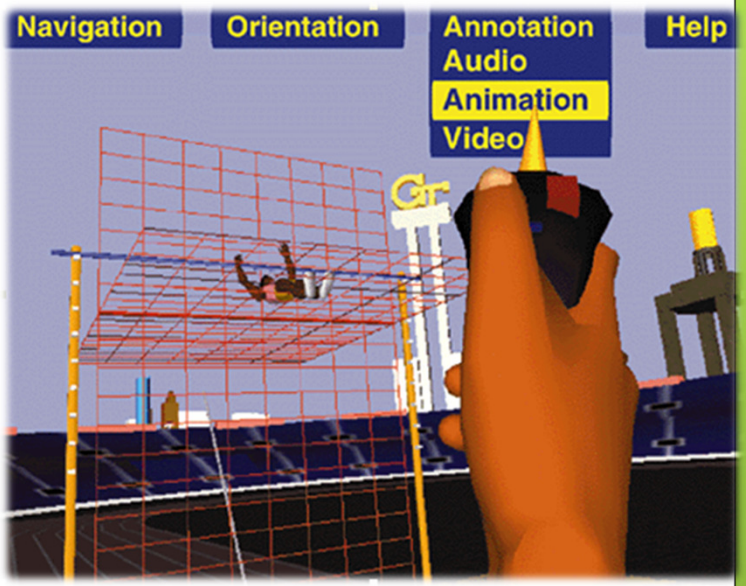
Inventing 3DUIs – Adapting from the Real World

- Adapt artifacts, ideas, philosophies, domains
- Architecture and movies
- Real world metaphors
- Examples
  - virtual vehicle
  - flashlight
  - shadows



Inventing 3DUIs – Adapting from 2D

- 2D UIs studied extensively
- Most people fluent with 2D interaction
- Can be easier than 3D
- Approaches
  - 2D overlay
  - Elements in 3D environment
  - 2D interaction with 3D objects
  - UI on separate device, e.g., Ipad



Inventing 3DUIs – Magic and Aesthetics

- Real power of 3DUIs
  - better reality
  - alternate reality
- Overcome human limitations
- Reduces effects of technological limitations

http://www.cantonmagicrafters.com/images/rabbit.jpg

Magic: Cultural Clichés & Metaphors

- Examples: Flying carpet, Go-Go, WIM
- Advantages:
  - easy to understand if you know the metaphor
  - usually they are very enjoyable
  - many metaphors are available
  - need not to be learned
- Disadvantages:
  - the metaphors can be misleading
  - the metaphors are often rooted in culture
  - it is difficult to come up with good magic metaphor

Magic: Violating Assumptions

- Can we systematically design and evaluate new interfaces by systematically violating our own assumptions? -- Jeff Pierce, CMU
  - Examples
    - what if 2 objects can occupy the same place in space and time?
    - what if we can make time go backwards?
    - what if we have a technology that has no flaws?
- Advantages:
  - systematic approach toward inventing 3D user interfaces
- Disadvantages
  - how far can we violate our assumptions?

3D UI Evaluation

User Evaluation in 3DUIs

- Was missing component for many years
  - novelty
  - limitless possibilities
  - exploration of design space
- Field has matured
  - Need to compare
    - devices
    - interaction techniques
    - applications
    - etc…

Purposes of Evaluation

- Evaluation – analysis, assessment, and testing of an artifact
- Problem identification and redesign
- General usability understanding
- Performance models

Some Terminology

- Usability – everything about an artifact and what affect a person's use of an artifact
- Evaluator – person who designs, administers, implements, or analyzes an evaluation
- Subject – person who takes part in the evaluation

Evaluation Tools

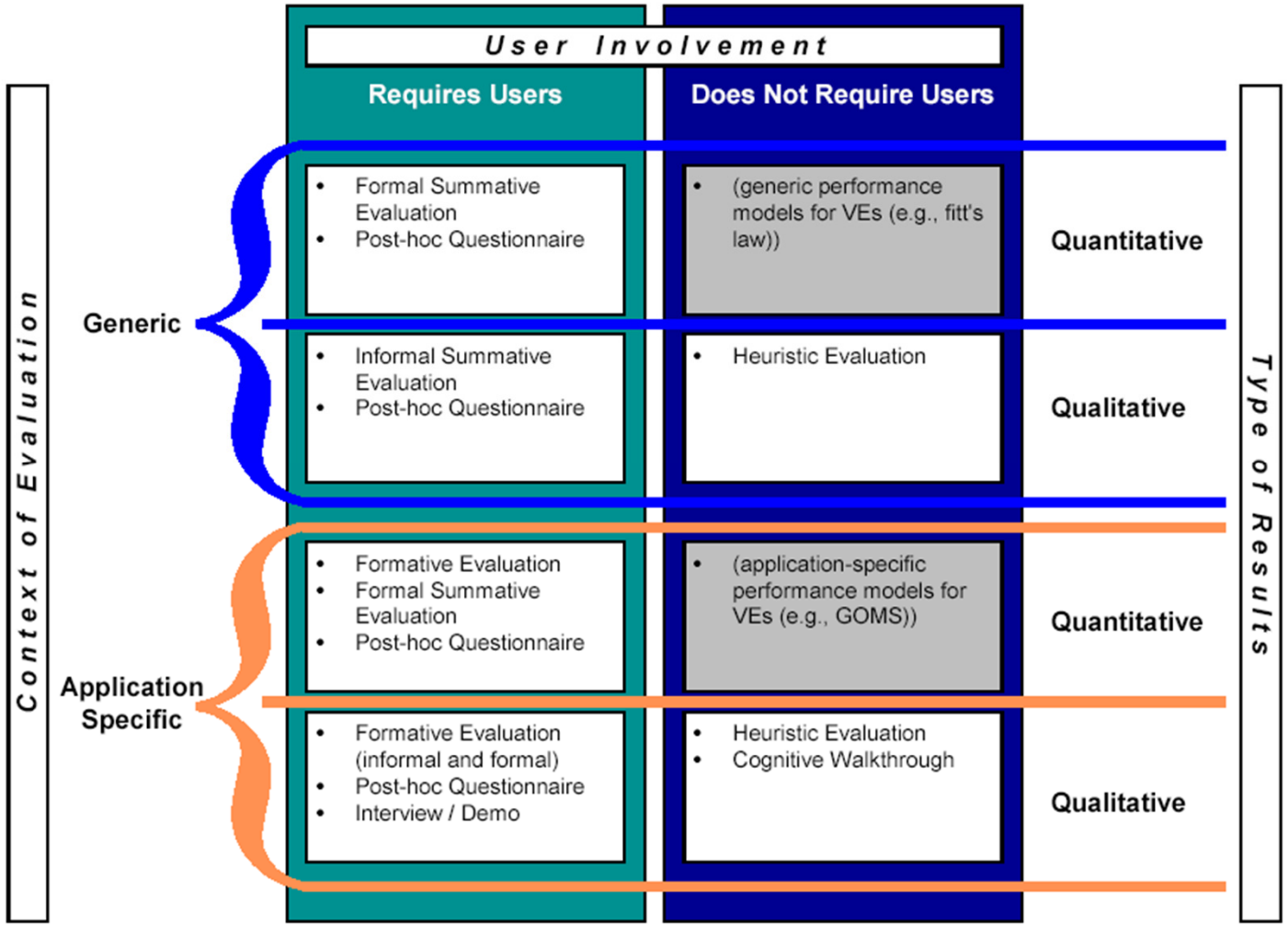
- User task analysis
  - generates list of detailed task descriptions, sequences, user work, and information flow
- Scenarios
  - built from task analysis
  - important for experiment design
- Taxonomy
  - science of classification
  - break techniques into components
  - used in evaluation process
- Prototyping
  - need to have something to test
  - paper-based sketches
  - Wizard of Oz approach

Evaluation Methods

- Cognitive walkthrough
- Heuristic evaluation
- Formative evaluation
  - observational user studies
  - questionnaires, interviews
- Summative evaluation
  - task-based usability evaluation
  - formal experimentation
- Questionnaires
- Interviews and Demos

Sequential evaluation

Testbed evaluation


User Involvement
Requires Users
Does Not Require Users
Context of Evaluation
Type of Results
Generic
Application Specific
Formal Summative Evaluation
Post-hoc Questionnaire
(generic performance models for VEs (e.g., fitt's law))
Quantitative
Informal Summative Evaluation
Post-hoc Questionnaire
Heuristic Evaluation
Qualitative
Formative Evaluation
Formal Summative Evaluation
Post-hoc Questionnaire
(application-specific performance models for VEs (e.g., GOMS))
Quantitative
Formative Evaluation (informal and formal)
Post-hoc Questionnaire
Interview / Demo
Heuristic Evaluation
Cognitive Walkthrough
Qualitative

Evaluation Metrics – System Performance

- System performance metrics
- Avg. frame rate (fps)
- Avg. latency / lag (msec)
- Variability in frame rate / lag
- Network delay
- Distortion

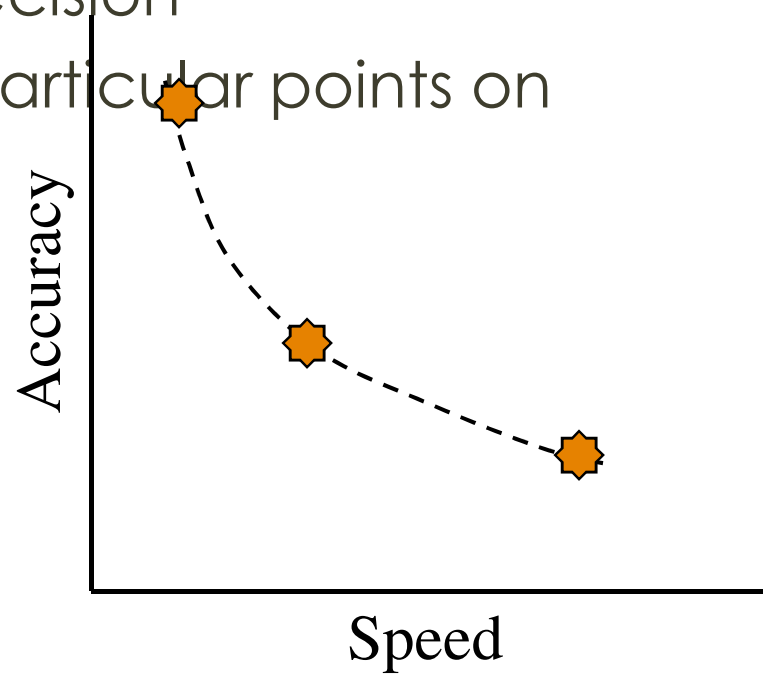
- Only important for its effects on user performance / preference
  - frame rate affects presence
  - net delay affects collaboration
- Necessary, but not sufficient

Evaluation Metrics – Task Performance

- Speed / efficiency
- Accuracy
- Domain-specific metrics
  - education: learning
  - training: spatial awareness
  - design: expressiveness

Speed-Accuracy Tradeoff

- Subjects will make a decision
- Must explicitly look at particular points on the curve
- *Manage* tradeoff



Evaluation Metrics – User Preference

- Ease of use / learning
- Presence
- User comfort
- Usually subjective (measured in questionnaires, interviews)

User Preference in the Interface

- UI goals
  - ease of use
  - ease of learning
  - affordances
  - unobtrusiveness
  - etc.
- Achieving these goals leads to *usability*
- Crucial for effective applications

User Comfort

- Simulator sickness
- Aftereffects of VE exposure
- Arm/hand strain
- Eye strain

Measuring User Comfort

- Rating scales
- Questionnaires
  - Kennedy - SSQ
- Objective measures
  - Stanney - measuring aftereffects

Characteristics of 3DUI Evaluation

- Physical environment
- Evaluator issues
- User issues
- Evaluation type issues
- Misc. issues

Physical Environment Issues

- Utilizes nontraditional input and output devices
- Many displace do not allow multiple simultaneous viewers
- Think-aloud and voice recognition
- Mobility and video recording
- Collaborative UIs and network behavior

Evaluator Issues

- May require more than one
- Breaking presence
- No evaluator intervention means robust software
  - instructions must be detailed
- Challenges with multimodal interfaces

User Issues

- Selection of subject pool
  - 3DUIs may not be well understood
- Novice vs. expert users
- Number of subjected needed may be larger than normal (novelty)
- Users must adapt to wide variety of situations
- Effects of cybersickness

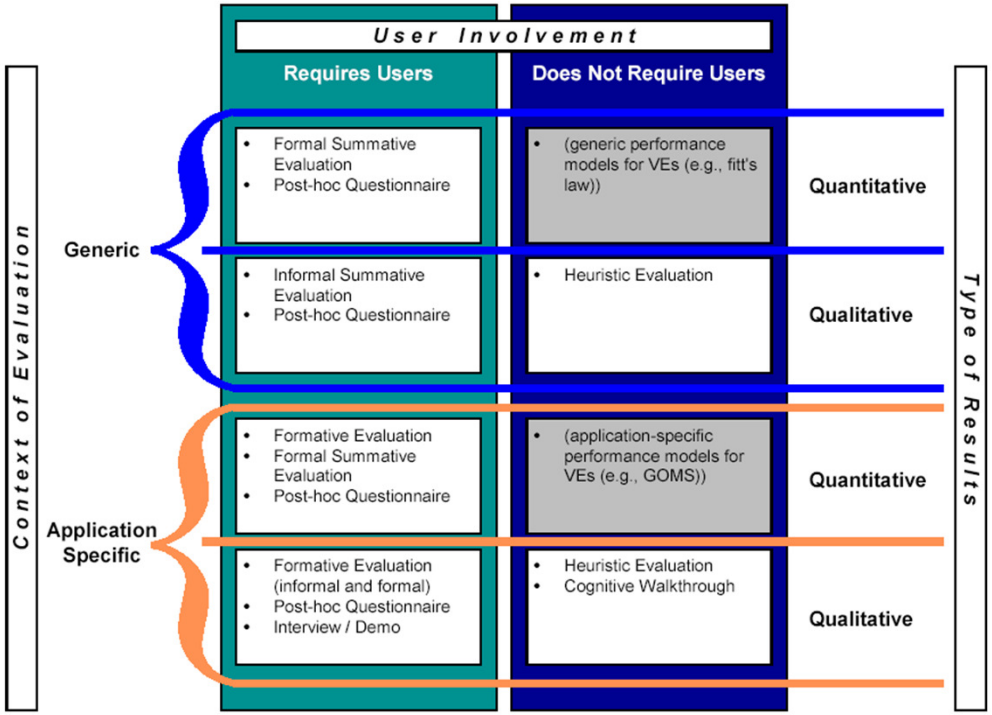
Evaluation Type Issues

- Heuristic evaluation difficult due to lack of guidelines
- Not many performance models for 3DUIs
- Automated tools are important
  - not many of them for 3DUIs
  - Multi-attribute Usability Evaluation Tool for Virtual Environments (MAUVE) – Stanney et al. 2000
- Statistical validity and 3DUI hardware
  - many factors to consider

Miscellaneous Issues

- Focus at a lower level
  - difficult to evaluate on application level
  - no set 3DUI standards
- Generalization of results

Usability Evaluation in 3DUIs



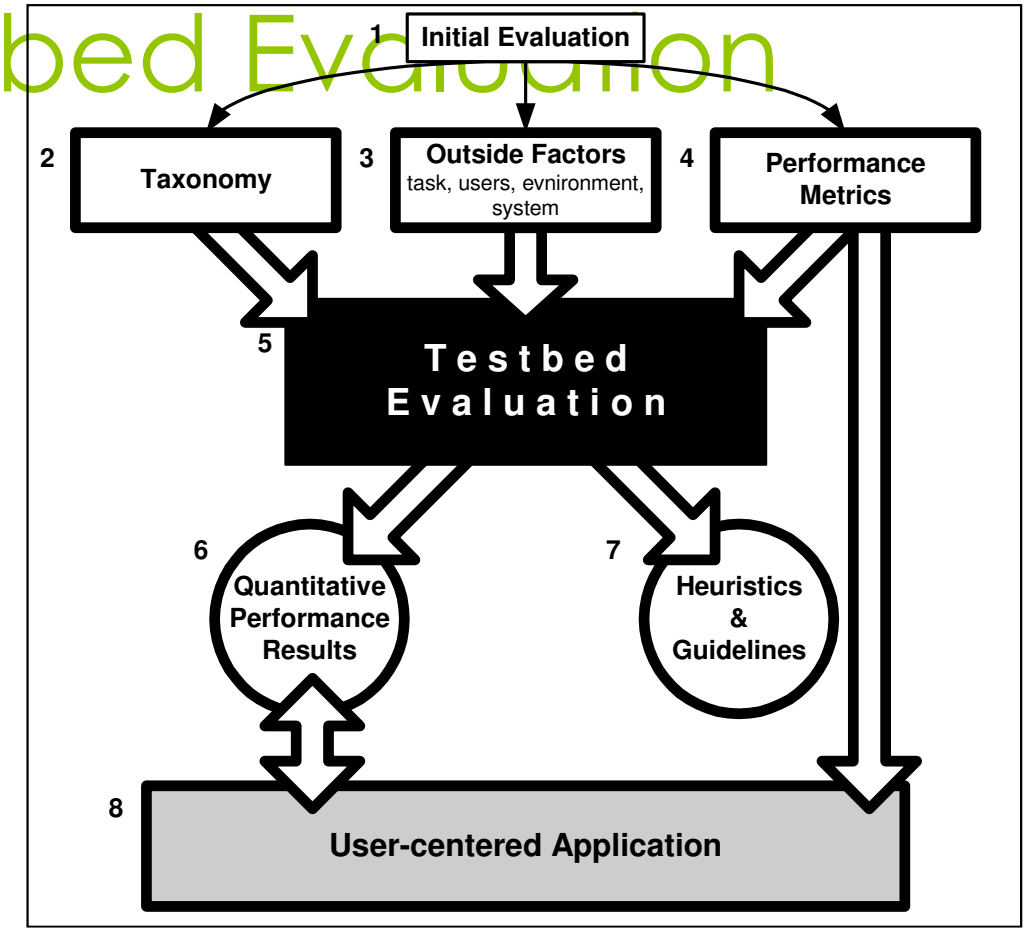
Classification Shortcoming

- 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

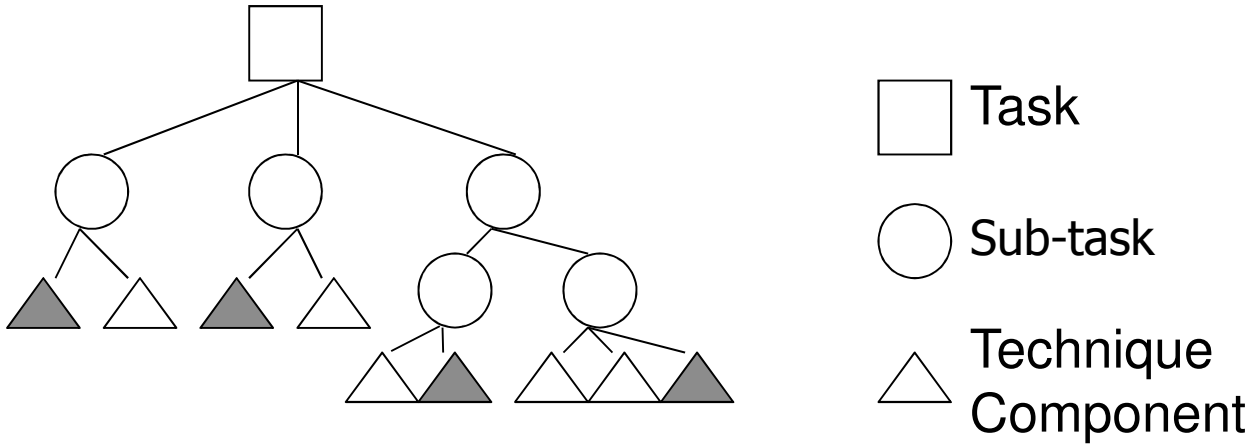


Testbed Evaluation – Initial Evaluation

- 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

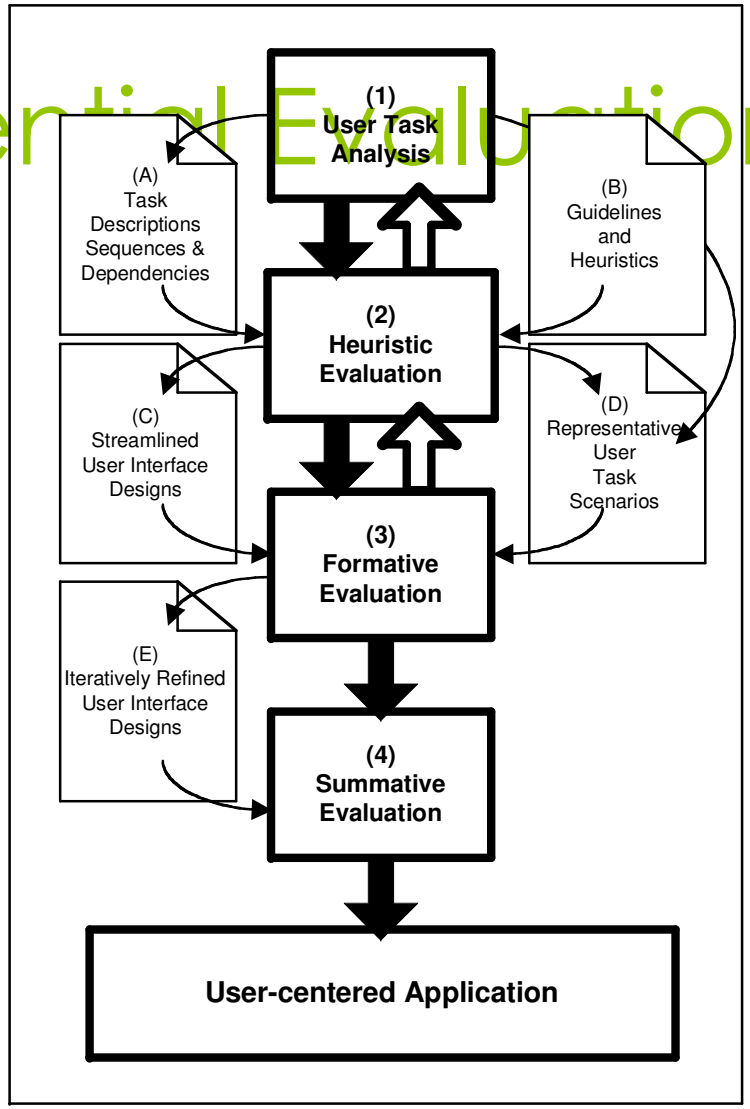
Testbed Evaluation Experiments

- 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 perform
  period
  - one to three users
  - two to three evaluators
  - Four cycles
- Guideline-based evaluatio
- 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

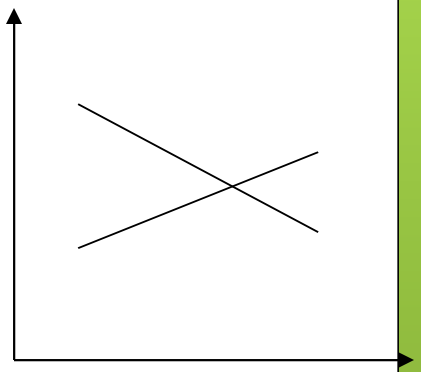
- Main variable of interest (e.g. interaction technique)
- Secondary variables
  - task characteristics
  - environment characteristics
  - system characteristics
  - user characteristics

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