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…
Introduction

Look at 3D UI designs in terms of user experience (including usability, usefulness, and emotional impact)

Must critically analyze, assess, and compare devices, interaction techniques, UIs, and applications

If 3D UIs are to be used in the real world

Purposes of Evaluation

Evaluation: analysis, assessment, and testing of an artifact

Usability-problem identification and UI redesign are the main goals of evaluation

General understanding gained from evaluation can lead to design guidelines

A more-ambitious goal of UI evaluation is the development of models that predict user performance
Introduction

Terminology

- **Evaluator**: a person who designs, implements, administers, or analyzes an evaluation
- **User (or participant)**: a person who takes part in an evaluation by using the interface, performing tasks, or answering questions
- **Evaluation method**: particular steps used in an evaluation
- **Evaluation approach**: a combination of methods, used in a particular sequence, to form a complete usability evaluation

Chapter Roadmap

- Evaluation methods for 3D UIs
- Evaluation metrics for 3D UIs
- Characteristics of 3D UI evaluations
- Classification of evaluation methods
- Three Multimethod Approaches
- Guidelines for 3D UI Evaluation
- Case Studies
Evaluation Methods for 3D UIs

- Cognitive walkthrough: stepping through common tasks that a user would perform and evaluating the interface's ability to support each step
- Heuristic evaluation: several usability experts separately evaluate a UI design by applying a set of design guidelines
- Formative evaluation: an observational, empirical evaluation that identifies usability problems by iteratively placing representative users in task-based scenarios

Summative evaluation:
- Comparing the usability of a UI to target usability values, or
- Comparing two or more UI designs, components, and/or techniques

Formal summative evaluations use:
- Research questions
- Independent variables (manipulated among multiple levels)
- Dependent variables
- Factorial designs and conditions
Evaluation Methods for 3D UIs

- Questionnaire: a set of questions used to obtain information from users before or after they have participated in an evaluation
- Interview: gathering information from users by talking directly to them

Evaluation Metrics for 3D UIs

**System Performance Metrics**
- Frame rate
- Latency
- Network delay
- Optical distortion
- Etc.
Evaluation Metrics for 3D UIs

Task Performance Metrics

- Speed
- Accuracy
- Errors

Subjective Response Metrics

- Presence: the “feeling of being there”
- Cybersickness: symptomatically similar to motion sickness and may result from mismatches in sensory information
- User comfort: strains on arms/hands/eyes
Characteristics of 3D UI Evaluations

**Physical Environment Issues**
- Evaluator must ensure that the user does not bump into physical objects, trip over cables, or move outside the tracking space
- Hardware or software must be set up so that the evaluator can see the same image as the user
- Think-aloud protocols are difficult to use with speech recognition as an interact technique
- Recording video of both the user and the interface is often difficult
- Collaborative 3D applications present several complications

**Evaluator Issues**
- Evaluators can cause breaks in presence if the user senses them
- Experimental applications should be robust enough that the evaluator does not have to interrupt the session to fix a problem
- Multiple evaluators may be needed due to the complexity of 3D UI hardware and software
- It is very difficult for an evaluator to observe multiple input streams, which are common to many 3D UIs, simultaneously and record an accurate log of the user’s actions
- Automated data collection is very important
Characteristics of 3D UI Evaluations

**User Issues**
- The target user population for a 3D application or interaction technique may not be known or well understood.
- It may be difficult to differentiate between novice and expert users because there are few potential participants who would be experts.
- The number of participants needed to obtain a good picture of performance may be larger than for traditional usability evaluations.
- Users must be able to adapt to a wide variety of situations for within-subject evaluations that compare two or more 3D UIs.
- 3D UI evaluations must consider the effects of cybersickness and fatigue.
- Presence is often required in VE evaluations.

**Evaluation Type Issues**
- Automated data collection of system and task performance metrics is nearly a necessity.
- Heuristic evaluations are very difficult due to a lack of verified guidelines for 3D UI design.
- Usability inspections are difficult to perform on early prototypes, because 3D UIs must be experienced firsthand.
- Few performance models have been developed for or adapted to 3D UIs.
- Statistical 3D UI experimental evaluations may be either overly complex or overly simplistic.
Characteristics of 3D UI Evaluations

General Issues

- 3D UI evaluations most often evaluate lower-level components, such as interaction techniques or input devices because there are no interface standards.
- It is important to report information about the apparatus with which the evaluation was performed and to evaluate with a range of setups if possible.
- It is the responsibility of 3D UI evaluators to ensure that the proper steps are taken to protect their human subjects.

Classification of Evaluation Methods

- Three key characteristics:
  - Involvement of representative users: participants required or not.
  - Context of evaluation: generic or application-specific context.
  - Types of results produced: qualitative or quantitative.
Classification of Evaluation Methods

Three Multimethod Approaches

Sequential Evaluation Approach
- Produces a usable and useful interface for a particular application
- Employs application-specific guidelines
- For domain-specific representative users
- Relies on application-specific user tasks
Three Multimethod Approaches

**Testbed Evaluation Approach**
- Empirically evaluates interaction techniques in a generic context
- Supported by a framework for design and evaluation
- Primarily aimed at researchers who are attempting to gain an in-depth understanding of interaction techniques and input devices

**Component Evaluation Approach**
- Focuses on the stages of action and the components that affect those stages
- The User-System Loop serves as the basis of the approach
- At each stage, there are components that affect the overall usability of the system
Three Multimethod Approaches

**Component Evaluation Approach**

- **Interaction Fidelity Components**
  - Interaction fidelity: objective degree of exactness with which real-world actions are reproduced in a 3D UI system
  - Biomechanical symmetry includes anthropometric symmetry, kinematic symmetry, and kinetic symmetry
  - Input veracity includes accuracy, precision, and latency
  - Control symmetry focuses on transfer function symmetry

- **Scenario Fidelity Components**
  - Scenario fidelity: objective degree of exactness with which behaviors, rules, and object properties are reproduced
  - Behaviors refer to artificial intelligence properties
  - Rules refer to physics and other models that determine what happen within the simulation
  - Object properties refer to dimensional and physics-related qualities of objects
Three Multimethod Approaches

**Component Evaluation Approach**

- Display Fidelity Components
  - Display fidelity: objective degree of exactness with which real-world sensory stimuli are reproduced by a system
  - Also referred to as immersion
  - Components of visual display fidelity include stereoscopy, field of view, field of regard, display resolution, display size, refresh rate, and frame rate

**Comparison of Approaches**

- What are the goals of the approach?
  - Sequential evaluation: iterate toward a better 3D UI
  - Testbed evaluation: finding generic performance characteristics of interaction techniques
  - Component evaluation: determining the main and interaction effects of specific system components for either an application-specific or generic context
Three Multimethod Approaches

Comparison of Approaches

When should the approach be used?
- Sequential evaluation: early and continually throughout the design cycle of a 3D application
- Testbed evaluation: before the design cycle begins
- Component evaluation: before the design cycle for knowledge of the general effects of one or more components or during the development of a 3D application to decide upon unclear design choices

In what situations is the approach useful?
- Sequential evaluation: throughout the design cycle of a 3D UI, but especially during the early stages
- Testbed evaluation: when choosing common interaction techniques and interface elements for a suite of applications
- Component evaluation: when making design choices that directly involve one or more system components
Three Multimethod Approaches

Comparison of Approaches

What are the costs of using the approach?
- Sequential evaluation: development of useful task scenarios and incorporating suggested design changes
- Testbed evaluation: very costly due to difficult experimental design and experiments requiring large numbers of trials
- Component evaluation: depends on whether employed for an application-specific or generic context

What are the benefits of using the approach?
- Sequential evaluation: likely to produce a more-refined and usable 3D UI
- Testbed evaluation: generality of the results
- Component evaluation: vary based on when and how the approach is used
Three Multimethod Approaches

Comparison of Approaches

- How are the approach’s results applied?
  - Sequential evaluation: results are tied directly to changes in the interface of the 3D application
  - Testbed evaluation: results are applicable to any 3D UI that uses the tasks studied with a testbed
  - Component evaluation: results are applicable to any 3D UI system that includes the system components evaluated

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

Guidelines for 3D UI Evaluation

**General Guidelines**

- Begin with informal evaluation.
- Acknowledge and plan for the differences between traditional UI and 3D UI evaluation.
- Choose an evaluation approach that meets your requirements.
- Use a wide range of metrics.
Guidelines for 3D UI Evaluation

Guidelines for Formal Experimentation
- Design experiments with general applicability.
- Use pilot studies to determine which variables should be tested in the main experiment.
- Use automated data collection for system performance and task performance metrics.
- Look for interactions between variables—rarely will a single technique be the best in all situations.

Case Studies

VR Gaming Case Study
- Working prototypes and iteration of individual interaction concepts and several rounds of iteration
- Prototype of the complete UI using just a couple of rooms representative of the entire game
- Key concepts:
  - Working prototypes are critical to understand the potential of 3D UI designs.
  - Be sure to evaluate the complete UI, not just the individual interaction techniques.
  - Start with usability evaluation, but for real 3D UI applications, go beyond usability to understand the broader user experience
Case Studies

Mobile AR Case Study
- Informal study with simple paper-based prototype
- Outdoor AR experiments were affected by lighting conditions
- Users experienced cognitive load and ergonomic issues
- Key concepts:
  - Be sure to evaluate AR systems in the environment in which the system is deployed.
  - Assess subjective mental load of more complex systems, as it may greatly affect performance.
  - Study ergonomics of systems that are used for lengthy time periods.

Conclusion
- Evaluation is almost always necessary
- Initial 3D UI design require assessment of usability and user experience so that the design can be iterated and improved
- Formal experimentation deepens our understanding of 3D interaction and provides new knowledge, guidelines, and models
Next Class

- 3DUI evaluation examples
- Readings
  - 3DUI Book - Chapter 11