


EECS 4441 Human-Computer Interaction

Topic #8: Evaluation – Part II

I. Scott MacKenzie

York University, Canada





Other Evaluation Methods (...other than a user study)

Observational Methods

- Field Studies
- Think Aloud
- Cooperative evaluation
- Protocol analysis
- Automated analysis
- Post-task walkthroughs

Field Studies

- Field studies are high in “relevance” (but low in “precision”)
 - Work studied in context
 - Real action is situated action
 - Physical and social environment both crucial
- Contrast with
 - Controlled experiment (aka user study)
 - Narrow/accurate results, ∴ high in “precision”
 - Environment is artificial, ∴ low in “relevance”

Observational Methods

- Field Studies
- Think Aloud
- Cooperative evaluation
- Protocol analysis
- Automated analysis
- Post-task walkthroughs



Think Aloud

- Users ...
 - Observed performing tasks
 - Describe out loud what he/she is doing and why
 - Describe what he/she thinks is happening, etc.
- Advantages
 - Simplicity - requires little expertise
 - Can provide useful insight
 - Can show how system is actually used
- Disadvantages
 - Subjective, selective, labour intensive
 - Act of describing may alter task performance
 - Answers are qualitative (hence inaccurate)

Observational Methods

- Field Studies
- Think Aloud
- Cooperative evaluation
- Protocol analysis
- Automated analysis
- Post-task walkthroughs



Cooperative Evaluation

- Variation on think aloud
- User and evaluator collaborate in evaluation
- Both user and evaluator can ask each other questions throughout
- Advantages
 - Less constrained and easier to use
 - User is encouraged to criticize system
 - Clarification possible
- Disadvantages
 - See previous slide
 - Act of discussing may alter task performance

Observational Methods

- Field Studies
- Think Aloud
- Cooperative evaluation
- Protocol analysis
- Automated analysis
- Post-task walkthroughs



Protocol Analysis

- Paper and pencil - cheap, limited to writing speed
- Plus videotaping and/or audio recording - accurate and realistic, needs special equipment, obtrusive
- Mixed use in practice
- Audio/video transcription difficult and requires skill
- Some automatic support tools available

Observational Methods

- Field Studies
- Think Aloud
- Cooperative evaluation
- Protocol analysis
- Automated analysis
- Post-task walkthroughs



Automated Analysis - EVA

- Videotaping with automatic annotation
- Post task walkthrough
 - Users react after interaction
 - Users indicate intention
- Advantages
 - Analyst/evaluator has time to focus on relevant incidents
 - Avoid excessive interruption of task
- Disadvantages
 - Lack of freshness
 - May be post-hoc interpretation of events
- E.g., Noldus *Observer XT* (next slide)

Noldus

ANIMAL BEHAVIOR

HUMAN BEHAVIOR

NOLDUS INNOVATIONWORKS

ABOUT NOLDUS

SUPPORT CENTER



Products ▾

Research areas ▾

Services ▾



THE OBSERVER® XT

The Observer XT is the professional and user-friendly software package for the collection, analysis, and presentation of observational data.

Request
Quote



Request
Free trial



<http://www.noldus.com>

Observational Methods

- Field Studies
- Think Aloud
- Cooperative evaluation
- Protocol analysis
- Automated analysis
- Post-task walkthroughs



Post-task Walkthroughs

- Transcript played back to participant for comment
 - Immediately → fresh in mind
 - Delayed → evaluator has time to identify questions
- Useful to identify reasons for actions and alternatives considered
- Necessary in cases where think aloud is not possible

Query Techniques

- Interviews
- Questionnaires



Interviews

- Analyst/evaluator questions user on one-to-one basis, usually based on prepared questions
- Informal, subjective and relatively cheap
- Advantages
 - Can be varied to suit context
 - Issues can be explored more fully
 - Can elicit user views and identify unanticipated problems
- Disadvantages
 - Very subjective
 - Time consuming

Query Techniques

- Interviews
- Questionnaires



Questionnaires

- Set of fixed questions given to users
- Advantages
 - Quick and reaches large user group
 - Can be analyzed more rigorously
- Disadvantages
 - Less flexible
 - Less probing

Questionnaires (continued)

- Need careful design
 - What information is required?
 - How are answers analyzed?
- Styles of question
 - General
 - Open-ended (e.g., *Can you suggest improvements to the interface?*)
- Ordinal
- Multiple choice
- Ranked

→ The interface was easy to understand
Disagree 1 2 3 4 5 Agree

Physiological Methods

- Eye tracking
- Physiological measurement



Eye Tracking

- Head or desk mounted equipment tracks the position of the eye
- Eye movement reflects the amount of cognitive processing a display requires
- Measurements include
 - Fixations
 - Eye maintains stable position.
 - Number and duration indicate level of difficulty with display
 - Saccades
 - Rapid eye movement from one point of interest to another
 - Scan paths
 - Moving straight to a target with a short fixation at the target is optimal (next slide)



Figure 1. On-screen keyboard and eye tracking device.

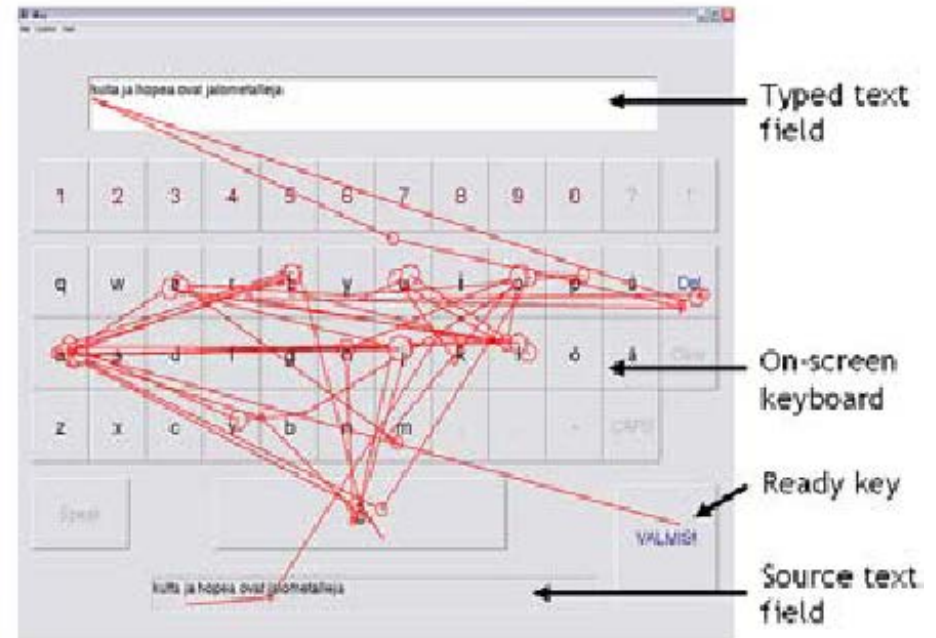


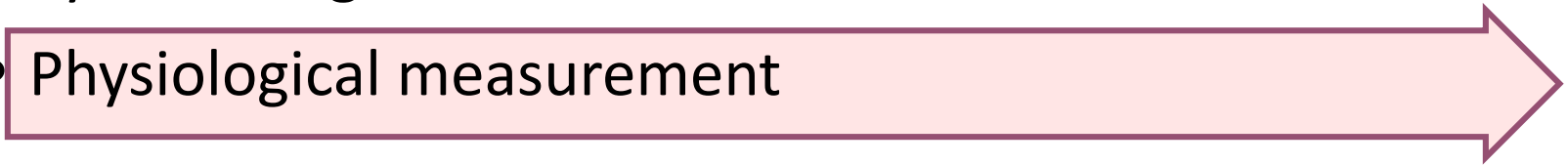
Figure 2. Experimental software and a gaze path of a user eye typing a phrase.

See...

Majaranta, P., MacKenzie, I. S., Aula, A., & Riih a, K.-J. (2003). Metrics for text entry research: Auditory and visual feedback during eye typing. *Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI 2003*, pp. 766-767. New York: ACM.

Physiological Methods

- Eye tracking
- Physiological measurement



Physiological Measurements

- Emotional response linked to physical changes
- These may help determine a user's reaction to an interface
- Measurements include:
 - Heart activity, including blood pressure, volume and pulse.
 - Activity of sweat glands: Galvanic Skin Response (GSR)
 - Electrical activity in muscle: electromyogram (EMG)
 - Electrical activity in brain: electroencephalogram (EEG)
- Some difficulty in interpreting these physiological responses - more research needed

Example

Using Affective State to Adapt Characters, NPCs, and the Environment in a First-Person Shooter Game

Faham Negini
Dept. of Computer Science
University of Saskatchewan
Saskatoon, Canada
faham.negini@usask.ca

Regan L. Mandryk
Dept. of Computer Science
University of Saskatchewan
Saskatoon, Canada
regan.mandryk@usask.ca

Kevin G. Stanley
Dept. of Computer Science
University of Saskatchewan
Saskatoon, Canada
kevin.stanley@usask.ca

IEEE-GEM 2014

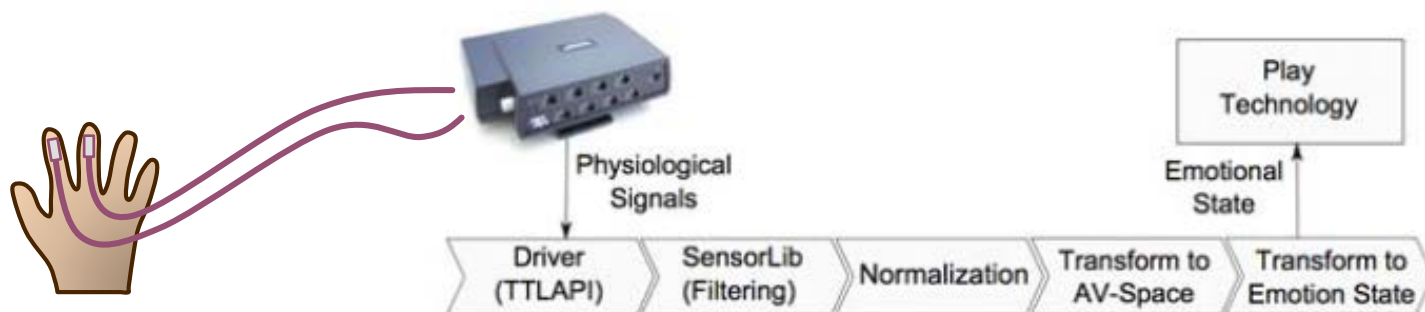


Fig. 2. A schematic of the affective engine modules.

Which Evaluation Method - Issues

- When: research, design, implementation
- Style of evaluation: laboratory vs. field
- How objective: subjective vs. objective
- Type of measures: qualitative vs. quantitative
- Level of information: high level vs. low level
- Level of interference: obtrusive vs. unobtrusive
- Resources available: time, subjects, equipment, expertise



Thank You