Subjective and Objective Measurement

Kai Römer

TU-München
Department of Informatics
Chair for Computer Aided Medical Procedures & Augmented Reality

30. April 2007

Introduction

- different types of user interfaces
- need for empirically test these user interfaces
- ability to improve the quality of user interfaces

Overview

$\Longrightarrow$ need for measuring usability.

- What is usability?
- How do I measure usability for time-critical user interfaces ?
  - subjectively
  - objectively
- How do I evaluate results of measurement?

What is usability?

- EN ISO 9241-11 European Standart for measuring usability
- a lot of prior art mainly based on ISO 9241-11

Defining Usability

The usability of a product is the degree to which specific users can achieve specific goals in a particular environment with effectiveness, efficiency and satisfaction. [1]

- really common definition based on ISO 9241-11
- seems not to be really helpfull for time-critical 3D UIs
- not mentioning the main issues of time-critical 3D UIs
- so new : not yet standardise.

Finding the Main Issues of Usability Testing

These problems should be in mind, when testing time-critical 3D UIs

- Information Overload (IO)
- Change Blindness (CB)
- Perceptual Tunneling (PT)
- Cognitive Capture (CC)

Measuring Distraction

- time-critical $\longrightarrow$ separation between main task and second task
- main task: driving, surgery, ...
- secondary task: using the interface
- interface may not disturb the main task!

Influence of usage of supportive interface on the main task needs to be measured.

Overview

- Objective Measurements
- Subjective Measurements

Objective Measurement

- task time
- eye tracking
- occlusion test
- quality of main task
- peripheral detection

Task Time

- time for application of secondary task
- main method for tests
- results:
  - $t_{secondarytask}$
  - ratio: $\frac{t_{secondarytask}}{t_{maintask}}$
- danger of distraction
- length of distraction

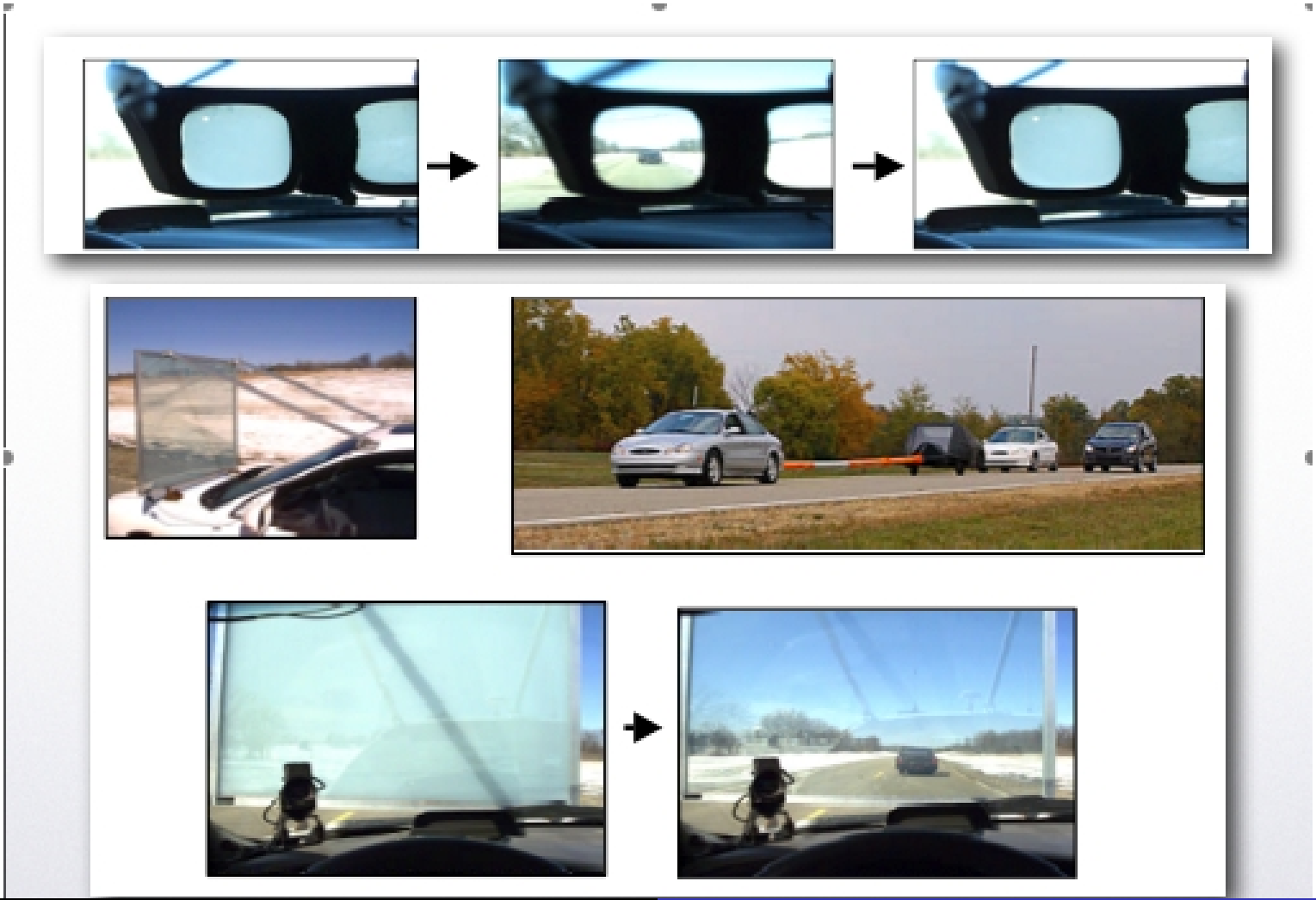
Eye Tracking: Glance Off Time

- time to
  - focus on an instrument +
  - read an instrument +
  - focus back on the main task
- time to mentally process information not included
- out of the loop effect
- ratio between time used for main task and secondary task

Occlusion Test

- measuring of time for secondary task
- main task neglected
- shutter glasses:
  - open for secondary task (1,5s)
  - closed for main task (about 4s)
- speed of perception, understanding, transcription
- interruptibility/resumability/time dependence
- problem: cheating by the test person

Occlusion Test



Quality of Main Task

- find distraction by measuring the quality of main task
- automotive
- track/speed keeping
- lane changing
- steering wheel reversal rate

Track/Speed Keeping Ability

- usage of interfaces
- measured in simulators or real cars
- cameras, tracking devices (GPS)
- cognitive capture, perceptual tunneling

Lane Change Task

- development by DaimlerChrysler
- car simulator
- road course with 3 lanes and indication, which lane to change to
- difference between driver's and reference trajectory
- parallel task execution
- testing stress and workload while performing additional tasks
- disadvantage: unrealistic, unnatural driving
- advantages: simple, quick, standardised

Lane Change Task

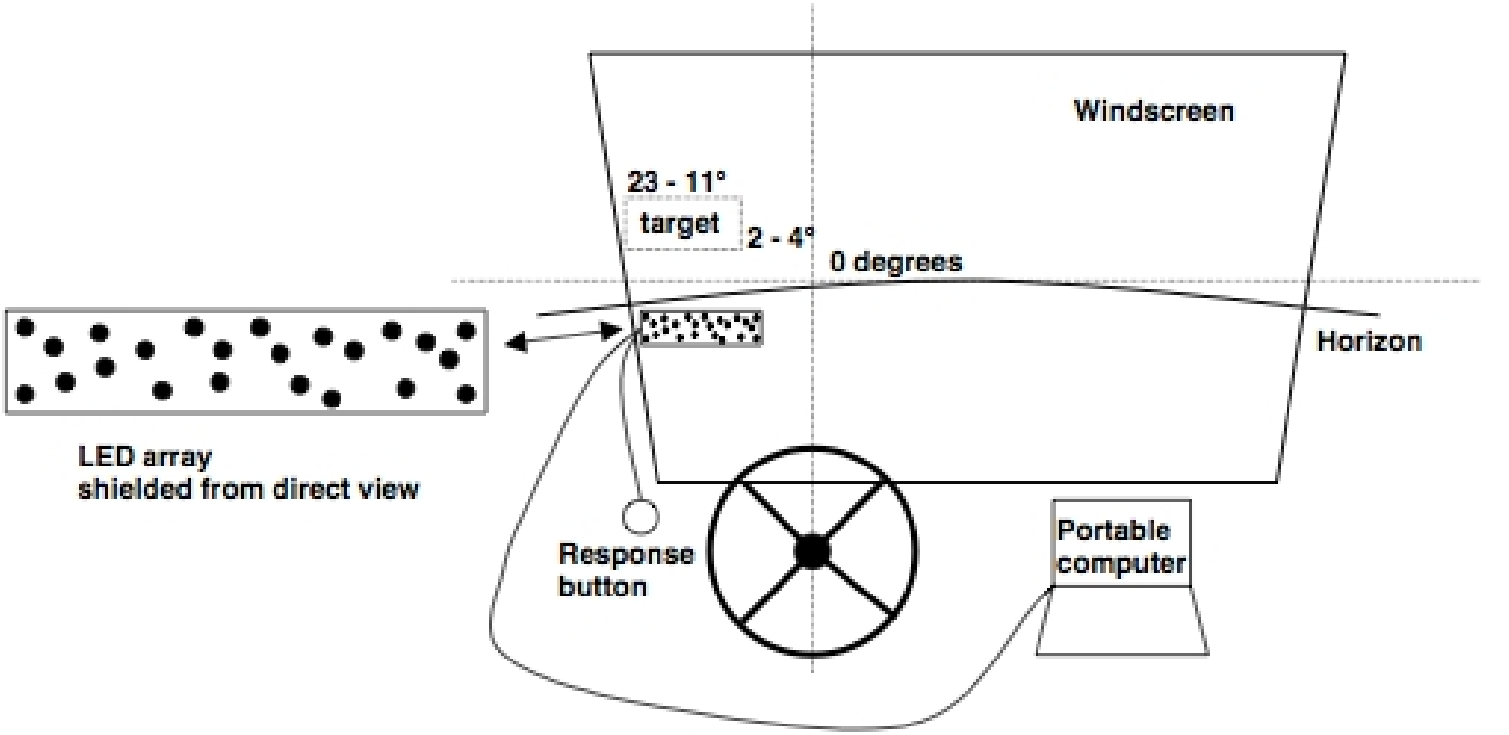
steering wheel reversal rate

- rate of intense steering wheel corrections
- usually after looking at a instrument
- peak detection algorithm, counted
- workload

Peripheral Detection Task

- extra task: respond to LED
- LEDs on horizontal line in different angels
- perceptual tunneling (errors)
- cognitive capture (reaction time)
- performance indices (workload) (average reaction time, errors)
- PDT is a concept, not a standardised test

Peripheral Detection Task

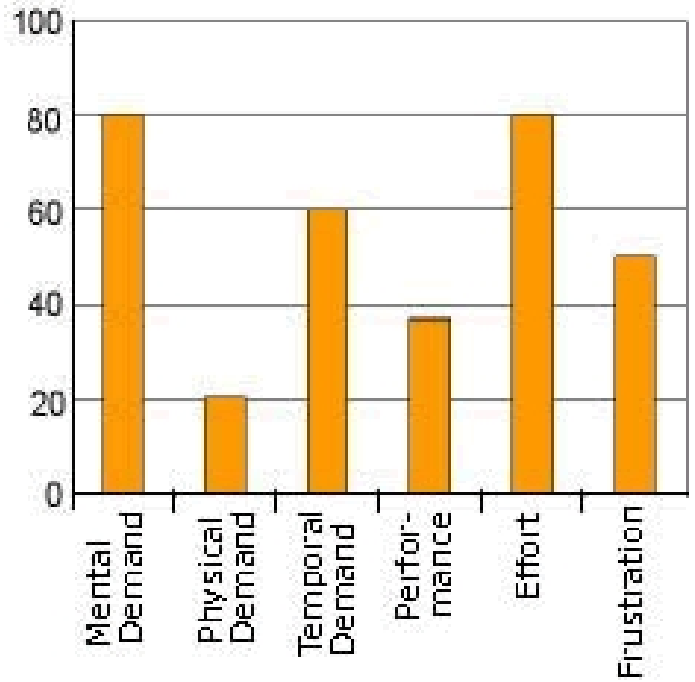


Subjective Measurements

- subjective measurement means asking the users
- several given questionnaires
- most common: NASA-TLX, SWAT

Different Questionairs

- NASA-TLX: NASA Task Load Index
  - mental demand
  - physical demand
  - temporal demand
  - performance
  - effort level
  - frustration level



Subjective Measurement

- SWAT: Subjective Workload Assessment Techniques
  - loads for time
  - mental effort
  - psychological stress

Questionaire

- Questionaire is answered directly after the test.
- advantages
  - ease of implementation
  - low cost
  - limited intrusion on task performance
- disadvantages: lack of precision due to repetition, understanding
- information overload

Cognitive Walkthrough

- test user tells what he is doing/feeling
- work flow test
- thought flow

Evaluation of Test Results

What does a Test look like?

- group of representative participants
- define structure of tests
- order of tests and participants
  - within subject
  - between subject
- what values should be measured
- evaluate the results

Goals

- Many numbers $\longrightarrow$ few numbers
- Measures of central tendency:
  - Mean: average
  - Median: middle data value
  - Mode most common data value
- Measures of variability / dispersion

Descriptive Statistics

- Mean:

$$\overline{X} = \frac{\sum X}{N}$$

- Mean absolute deviation:

$$\frac{\sum |X - \overline{X}|}{N} = MS$$

Descriptive Statistics

- Variance:

$$s^2 = \frac{\sum(X - \overline{X})^2}{N - 1}$$

- Standard deviation:

$$s = \sqrt{\frac{\sum(X - \overline{X})^2}{N - 1}}$$

Hypothesis Testing

Goal: conclude characteristics from samples

- Is system A significantly better than System B?
- Is there any siginigcant difference anyway?

Testing Hypothesises

1. hypothesis
2. develop testable hypothesis H_1
3. develop null hypothesis (logical opposite) H_0
4. calculate:

$$H_0 : p(X|H_0)$$

5. proves: Probability of logical opposite is small!
6. common significance levels are: $\alpha <= 0.05; \alpha <= 0.01$

Methodes

- two variances: t-test
- several variances ANOVA

t-Test

The t-test tells us, if the variation between two groups is "significant,,.

- 2 pairs, t-distribution
- implemented in e.g. Excel, SPSS
- n_1,n_2: sample count
- $\overline{y}_1$,$\overline{y}_2$: mean value
- s: mean variance
- compare to value table of distribution

$$t = \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} * \frac{\overline{y}_1 - \overline{y}_2}{s}$$

ANOVA

- analysis of variance
- 2 or more groups
- f-distribution

$$F = \frac{n_1 n_2 (\bar{x}_1 - \bar{x}_2)^2}{(n_1 + n_2) var_g}$$

Summary

- introduced several surveys of testing 3D UIs in time-critical environments
- presented, how tests can deliver reliable results
- showed ways of processing gathered data

Thank you for your attention!

 NPL Report DICT 163/90: A preliminary design for a methodology for experimentally measuring usability, R E Renger, 1990

 http://www.hf.faa.gov/Webtraining/

Distraction from main Task

- main issue: distraction from main task.
- information overload (IO)
  - too much information
  - large amount of information
  - high rate of new information
  - contradictions in available information
  - low signal to nosie ratio
  - problem to remain informed
  - problem to make a decision

Distraction from main Task

- perceptual tunneling (PT)
  - fucusing on one stimulus like flashing signal
  - leading to neglection of main task
- cognitive capture (CC)
  - lost in thought
  - caused by e.g. instruments requireing highly cognitive involvment
  - leads to a loss of situational awareness

Situation Awareness

clear understanding of:

- what is going on in the environment
- what is going to happen in the nearest future