Introduction

Tanja Schultz,
Christoph Amma & Felix Putze
19.10.2012

„Design and Evaluation of innovative User Interfaces“
WS 2012/13
General Information: Lecture

- Lecture in the Hauptdiplom or Bachelor/Master
  - No special previous knowledge required
  - Every year in the Winter Semester, 2 SWS or 3 LP points

- Exam:
  - Diplom: Possible in „Kognitive Systeme“ and „Anthropomatik“
  - Bachelor: atomares Modul
  - Master:
    - Biosignalverarbeitung (9LP)
    - Sprachverarbeitung (9LP)
    - Menschliches Verhalten in der Interaktion (6LP)
    - Multimodale Mensch-Maschine Interaktion (9LP)
  - Exam needs to be done during the lecture time (Vorlesungszeit) and you need to register a date early
General Information: Lecture

- Dates:
  - Friday, 11:30 – 13:00
  - Building 50.34, Basement, Room -102
  - Start 19.10.2011 – End 08.02.2011

- Lecturers:
  - (Prof. Dr. Tanja Schultz)
  - Dipl.-Inform. Felix Putze
  - Dipl.-Inform. Christoph Amma
You can find all material related to the lecture under:

- [http://csl.anthropomatik.kit.edu](http://csl.anthropomatik.kit.edu)
- Lecture Slides in PDF format (password protected)
- Announcements, Corrections, Changes, Syllabus
- All papers referenced in the lecture (password protected)
  - Main papers will be available in advance of the corresponding lecture to give you the opportunity to prepare yourself for the lecture

Basis for exams:

- Content of the lecture (based on the slides)
- Main papers (will be marked in the slides)

Questions, Problems, Comments are always welcome:

- tanja.schultz@kit.edu
- felix.putze@kit.edu
- christoph.amma@kit.edu
Design und Evaluation von Innovativen Benutzeroberflächen

Kognitive Modellierung

Methoden der Biosignalverarbeitung

Multilinguale Mensch-Maschine-Kommunikation

Design und Evaluation innovativer Benutzeroberflächen

Biosignale und Benutzeroberflächen

Biosignalspezifische Multilinguale Mensch-Maschine-Kommunikation

CSL: Teaching
interACT:

Seit 2004 gemeinsames Forschungszentrum von Carnegie Mellon University (CMU) und Universität Karlsruhe (TH)

Februar 2007: Erweiterung um Hong Kong University of Science and Technology (HKUST)

September 2008: Erweiterung um Waseda Universität, Japan

Gemeinsame Forschungsprojekte zwischen CMU, UKA, HKUST und Waseda - Gastvorlesungen, Workshops, Sommerakademien, Studien- und Forschungsaufenthalte

Begrenzte Zahl von Stipendien für herausragende Studierende, die ihre Studien der Informatik mit einem Forschungsaufenthalt im Ausland verbinden wollen!
interACT Stipendien

- Wer wird gefördert?
  - Studierende der Informatik und Informationswirtschaft der Universität Karlsruhe (TH), der Carnegie Mellon University, der HKUST und der Waseda University

- Was wird gefördert?
  - Studien- & Bachelorarbeiten, bis zu 3 Monaten
  - Diplom- & Masterarbeiten, bis zu 8 Monaten
  - Doktorarbeiten, bis zu 12 Monaten

- Bewerbungsvoraussetzungen
  - Gute Englisch-Kenntnisse
  - Betreuender Professor an UKA und der Gastuniversität

Bewerbungsfristen und Information:
Frau Rödder, roedder@ira.uka.de
Introduction and Motivation
Classic user interfaces

- Unintuitive
- Slow
- Inconvenient
- Known interaction paradigms
- Established development process
Innovative user interfaces

- Anthropomatics: Symbiosis of man and machine
- Design user interfaces that make man-machine interaction...
  - More intuitive
  - More natural
  - More efficient
  - More robust
  - More joyful
- Make use of innovative input techniques:
  - Speech
  - Vision
  - Biosignals
- Interaction paradigms much less known
- No established design and evaluation process
- No standardized toolboxes
Examples: Clippy

- Integrated into Microsoft Office as assistant
- Monitors user’s behavior and tries to detect problematic situations and the user’s goal
- One of the most hated interface components of all time
- Possible Reasons:
  - Cognitive dissonance
  - Lacking social etiquette
  - Loss of control
Examples: Mobile Interfaces

- Modern Smartphones support very complex applications
  - Requirement for more elaborate input mechanisms
- Integration of new means of input for small, mobile devices
  - Gestures (on the screen, with the whole device)
  - Speech
- Techniques known to work in controlled environments
  - In the lab, in the office
  - With experienced users
- Now: Mobile application, mass of users → new surprises
  - Many changing (sometimes adverse) environments
  - Large and diverse user base (from kids to elderly)
  - No control over usage patterns
Example Systems: Gesture based Interfaces

- Usage of arm and finger motions as natural input modality
  - No extra device
  - Allows 3D interaction
  - Interaction with computers ≈ interaction with everyday objects

- MIT 6th sense
  - Camera and beamer worn like a necklace
  - Beamer allows to use everyday environment as screen

Pictures: MIT 6th Sense project
Examples: Adaptive Avatar

- "Virtual Co-driver" gives useful and entertaining information
- Car offers challenging scenario
  - Ever-changing environment
  - Changing inner states of the driver during long ride
  - Interaction is secondary task to driving
- Interface to measure the driver’s mental workload
- Cognitive model tracks what is going on in the user’s mind
- Under development at the Cognitive Systems Lab

How is this restaurant called?
It is called "Revolving UFO Restaurant"
Example: Humanoid Robots

- Offer interface to issue commands
  - Speech
  - Gestures
- May implicitly "guess" the user's intention
  - Observing behavioral patterns
  - Learning
- Goal: Permanent integration of humanoid robot in user's household
- Should behave socially adequate
  - Recognize the user's emotional state
  - Adapt to the user's personality
  - Adapt to the user's expertise
  - Express their one "internal state"
Examples: Smart Homes

- A network of small intelligent devices forms a smart home
  - Automatically adjusts volume of music when phone rings
  - Prepares breakfast just before you wake up
  - Reminds you when you run out of milk (or orders it itself)
- Ambient Intelligence: Ubiquitous, invisible, smart components
- Smart home learns preferences and requirements of users
- Reduce amount of explicit commands
- How little control do we actually find comfortable?
- How does this development influence our personal lives?
Researchers and developers from many fields are interested in man-machine interaction

- Computer Scientists: How to design intelligent algorithms and interfaces?
- Product Designer: How to translate academic findings into products that survive a field test
- Ergonomists: How do new interfaces impact our work and efficiency?
- Statisticians: How to generate reliable quantitative results in evaluations?
- Psychologists: How do we perceive machines with such interfaces?
- Social Scientists: How does society influence the reception of new interfaces (and vice versa)?
- Philosophers: What are the implications of “intelligent” devices?

→ Expertise from one field is not enough
Design of innovative user interfaces

- How to incorporate new input paradigms (e.g. gesture input) into existing interfaces?
- How to combine multiple modalities?
- How to define strategies to handle the user’s input?
- How to transfer a system from the lab to the field?
- How to make systems survive reality?
Evaluation of innovative user interfaces

• What do people expect from an innovative user interface?
• How do people behave when confronted with an innovative user interface?
• Can we measure how good a new user interface is?
• Can we measure if it is actually better than the old one?
• What does “good” actually mean in this context?
Paper-driven Lecture

• Structure of the lecture: Guided by seminal papers
  • Selection of important publications in journals and conference proceedings
  • Present relevant finding and results
  • Sound methodology of general interest

• One or two subsequent lectures will be centered around one paper, covering both content and methodology

• Material from other papers as supplement

• Main papers (≈6) mandatory reading for exam

• Why no book? → There is none that...
  • is up to date in the described techniques
  • contains multiple viewpoints
  • covers the basics clear enough
Topics for the semester

- Multimodal Interaction
- Computers as Social Actors
- Elementary Statistics
- Emotional Interfaces
- Explicit vs. Implicit Interfaces
- Usability
- Fusion
User Interface: Definition

- A user interface is the system or infrastructure by which people interact with a machine
  - Consists of hardware (input device) and software (input processing)
  - Input: User commands, context information
  - Output: System output, feedback
- Innovative user interface: user interface which makes use of novel techniques to achieve a more intuitive interaction
- Interaction paradigm: „schools of thought“, i.e. general conception of how man-machine interaction takes place
  - Computer as tool: fixed set of commands, defined outcome
  - Computer as android: interact with computer = interact with human
Interaction Paradigms over Time

• First computers:
  • Only operable by experts
  • Direct manipulation of system components

• Command era:
  • Issue commands on command line
  • Still in use for certain tasks
  • Too cryptic for naive users

• Metaphor era:
  • Graphical user interfaces
  • Metaphor for the system, e.g.: Desktop (with Trash, Folders, ...)
  • Mouse as main manipulator

• Intuitive era:
  • Innovative input modalities
  • Interaction becomes more natural
How do we design interfaces?

- Implement known interaction paradigms...
  - Windows
  - Buttons
  - Menus
  - Mouse cursor

- ...using standardized development toolkits
  - Provide ready-to-use functionality to employ those paradigms
  - Often even integrated into the user’s operating system

Qt GUI framework (Trolltech)
Innovation in Interface Design

- Interface design requires constant refinement
- Established design tools and guidelines may become obsolete
  - Tasks change
    - mobile applications
    - 24/7 applications
  - Users change
    - get used to electronics and gadgets
    - gentrification
  - Technologies change
    - gesture and speech recognition ready for deployment
    - more upcoming technology (e.g. emotion recognition)
Basic building blocks of an innovative user interface

Human

Modalities

- Speech
- Motion
- Haptics
- Inner State
- Vision

Interaction System

- Interpretation
- Generation
- Strategy
This introduction should give you a rough overview on:
• The typical building blocks of interaction systems
• Common pitfalls you need to be aware of
• A basic idea of what happens in every step and why
## Modalities and Signals

- Different modalities require different signals to measure
- Often there is more than one possibility

<table>
<thead>
<tr>
<th>Modalities</th>
<th>Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech:</td>
<td>Acoustic, (EMG, Vision, Ultrasound)</td>
</tr>
<tr>
<td>Motion, Pose:</td>
<td>Vision, Acceleration, EMG</td>
</tr>
<tr>
<td>Haptics:</td>
<td>Pressure Sensors, Force Sensors</td>
</tr>
<tr>
<td>Gaze:</td>
<td>Vision, EOG</td>
</tr>
<tr>
<td>Inner States:</td>
<td>EEG, EDA, Vision, Acoustic</td>
</tr>
</tbody>
</table>

- Typical selection criteria:
  - Accuracy
  - Reliability
  - Practicability
  - Mobility
Signal Acquisition

- **Measure the right thing!**
  - There is no way to correct bad or wrong sensor signals
  - Choose the right hardware sensor
  - Choose correct placement

- **Transform the analog signal to a digital signal**

![Diagram of signal acquisition](image)

- **Signal quantization (A/D Converter)**
  - Transform analog signals to digital signals
  - Continuous to discrete conversion

**Notes:**
- Measure the right thing!
- Choose correct placement
- Choose the right hardware sensor
- There is no way to correct bad or wrong sensor signals
Signal Acquisition

Problems with raw signals from sensors of any kind:

- **Sensor errors**
  - Sensor quality dependant (webcam vs. high-end cam)
  → Solution: Use a better system

- **Artifacts, noise, outliers:**
  - Muscle activity in the EEG signal (eye blinks)
  → Solution: use specialized algorithms to remove artifacts

- **Sensor configuration dependent:**
  - Accelerometers are temperature dependent
  → Solution: Calibrate sensor, normalize data

→ Data needs preprocessing
Imagine a gestural interface to move slides by hand with two modes discriminated by speed of movement (normal/fast forward). The sensor supplies 3d coordinates.

- Is the resulting signal well suited to discriminate both modes?
  - Plot shows only one dimension

- Amplitude reveals no Information about mode
  - Example: First Derivate from the positional coordinates
Feature Extraction

- Example: From signals to sequences of feature vectors

**Framing**

**Feature Extraction**

- Derivate
- Mean
- Feature Vectors

\[
\begin{align*}
\frac{dx}{dy} & \\
\frac{dy}{dz} & \\
\frac{dz}{x} & \\
\frac{d}{y} & \\
\frac{d}{z} &
\end{align*}
\]

Stacking

\[
\begin{pmatrix}
\frac{dx}{dy} \\
\frac{dy}{dz} \\
\frac{dz}{x} \\
\frac{d}{y} \\
\frac{d}{z}
\end{pmatrix}
\]

\[
\begin{pmatrix}
\frac{dx_1}{dy_1} \\
\frac{dy_1}{dz_1} \\
\frac{dz_1}{x_1} \\
\frac{dx_2}{dy_2} \\
\frac{dy_2}{dz_2} \\
\frac{dz_2}{x_2} \\
\vdots
\end{pmatrix}
\]

\[
\begin{pmatrix}
\frac{dx_n}{dy_n} \\
\frac{dy_n}{dz_n} \\
\frac{dz_n}{x_n} \\
\frac{dx}{y} \\
\frac{d}{z}
\end{pmatrix}
\]

Classification of feature vectors
Pattern Recognition

- The signals need to be interpreted
- Patterns must be searched in the signal

Assign recorded data to

- Finite number of discrete categories
  - Classification Task
    - Example: Handwritten character recognition
- One or more continuous output variables
  - Regression Task
    - Example: Speed in gestural scrolling interface
- There are numerous different approaches and methods
Humans have large variabilities in their actions

- **Intrapersonal** differences:
  - We are not able to reproduce any kind of signal we emit a second time in exactly the same way

- **Interpersonal** differences:
  - Everybody writes, talks, moves, behaves differently

\[ \rightarrow \text{Interpersonal differences are much bigger than intrapersonal} \]

\[ \rightarrow \text{This requires advanced modeling techniques} \]

\[ \rightarrow \text{Pattern Recognition and Machine Learning techniques are necessary tools for interface developers} \]
Pattern Recognition

- We will have a look on a simple example of a statistical, parametric, supervised classification method.

- **Statistical:**
  - Learns model implicitly from examples
  - Usually needs lots of training data
  - Adapts well to real world data

- **Unsupervised**
  - Classes are known
  - Data is labeled

- **Parametric**
  - Assume probability distribution
  - Estimate parameters from data
Pattern Recognition: Example

- An imaginary Augmented Reality interface
- Inspired by the Glabelar Fader (Nakamura and Miyashita ‘10)
  - How to switch Augmented Reality Glasses on and off?
  - It might be intuitive to use the Glabella muscle
  - Augmentation is switched by frowning
  - Muscle Activity is measured via surface EMG

- Let’s assume we already have extracted features:
  - EMG Signal $\rightarrow$ Filtered $\rightarrow$ Power over time
  - We made some experiments and labeled the EMG data
  - 2 classes: ON / OFF
  - Histogram:
    - Samples of one class are approximately normally distributed
Pattern Recognition: Gaussian Classifier

- **Gaussian/Normal Distribution**
  - Probability Density Function:
    \[
    N(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (x - \mu)^2\right]
    \]

- **Classification**
  - One Gaussian for every class
  - Evaluate PDFs of all classes
  - Choose the \text{argmax}

- **How to find the parameters?**
  - Maximum-likelihood approach
  - Choose parameters in a way that:
    The Probability of the training data is maximized
Pattern Recognition: Challenges

- The problem of overfitting models
  - If the model has too much parameters
  - If we do not have realistic training data with high variability

- The problem of training data
  - „There is no data like more data!“
  - Realistic training data is often hard to collect
    - Must cover all real world situations
    - Typically involves recordings with lots of people
  - Training data must be labeled
    - If done manually, this is very costly
Design and Innovation

- Work with innovative user interfaces = constant improvement of components and concepts
- Innovation cycle:

  - Iterative process of two steps:
    - Design of a new or improved user interface
    - Evaluation of the interaction
Minimal Process Chain

- **Formative evaluation**
  - Before or during the design of an interface
  - Find problems in existing interaction patterns
  - Identify paths for improving the system

- **Design**
  - Technical realization of the system
  - Makes uses of several innovative interface components

- **Summative evaluation**
  - Does the new system improve over the old baseline in a technical sense?
  - Does the new system create a better user experience?

- *When the cook tastes the soup, that’s formative; When the guests taste the soup, that’s summative.* (Robert Stakes)
• Typical research questions/tasks:
  - Find the bottleneck of an existing user interface
  - Find out certain characteristics of human-human interaction
  - Find out how users react to certain system behavior

• Overall goal: Identify strategies for interface improvement

• Process guided by scientific hypotheses
  - Example: “Does the addition of a new modality improve efficiency?”

• Often also explorative component
  - Example: “What happens if we let the system mimic the user’s facial expression? How will users react?”
  - Leads to the formulation of new hypotheses
Methods of Formative Evaluation

- **Study theory of interaction**
  - Large corpus of literature on interaction
  - Provides a general model and design guidelines
  - Often backed up by many experiments
  - May not fit the specific requirements/constraints of the current task

- **Ask design experts**
  - Have experts review the specific question
  - Cognitive Walkthrough: role play certain use cases
  - May miss some effects or issues

- **Collect data**
  - Empirical approach
  - Include real user in the development process
  - May yield new and surprising findings
Theory of Interaction: Example

- „Magic 7“ (Miller, 1956): People can maintain seven items in their short time memory
- Many researchers drew conclusions for interface design
  - Not more than seven links on a web page
  - Not more than seven choices in a menu
- However, this generalization of a very specific psychological experiment is dangerous
  - Newer psychological evidence yields a much more diverse picture
  - Transfer to interface design is vague at best
  - Effect of memory on perception and interaction is not straight-forward
  - Leads to over-simplification
Design Guidelines: Example

1) Visibility of system status
2) Match between system and the real world
3) User control and freedom
4) Consistency and standards
5) Error prevention
6) Recognition rather than recall
7) Flexibility and efficiency of use
8) Aesthetic and minimalist design
9) Help users recognize, diagnose, and recover from errors
10) Help and documentation

Source: *10 heuristics in user interface design* (Jakob Nielsen)
Types of User Studies

- **Descriptive (Observational)**
  - Describe the state of affairs
  - No intervention, isolated study of different parameters
  - Example: How many people use the speech interface of their smartphone?

- **Relational**
  - Study the (undirected) relation between two measures
  - Does parameter A change if parameter B does?
  - Example: Is the judgment of a user interface related to the age of the subjects?

- **Causal**
  - Is the change of a dependent measure caused by another parameter?
  - Does a treatment (e.g. a change in the interface) have an effect?
  - Example: Does the addition of a new input modality improve system efficiency?
Empirical Measurement Methods

- **Observational:** Observe the user during interaction
  - Think aloud
  - Post-talk walkthrough
  - Analysis of recordings or transcriptions
- **Queries:** Ask the user on interaction behavior/preferences
  - Surveys
  - Self-report diaries
  - Focus groups
- **Both techniques are explorative and give qualitative results**
- **Experiments**
  - Follows certain scientific standards
  - Yields quantitative results
  - Tries to single out one specific effect
  - Only way to investigate causal research questions
Experiment: Example

- Hypothesis: „Faster reaction time of the interface increases user satisfaction“
- Form two groups of participants:
  - Group A interacts with a slow system
  - Group B interacts with a fast system
- Afterwards, measure user satisfaction (→ questionnaire)
- Critically assess the accordance between the hypothesis and the collected data
Subjects and Degree of Control

- Experiments can take place in many different forms
- Who are the users in the study?
  - Experts (e.g. the developers or their colleagues)
  - Naive recruited subjects (e.g. paid students)
  - Real users (of a deployed system)
- Different groups have different incentives of participating!
- How controlled is the user study?
  - In the lab
    - Controllable, reproducible conditions
    - No interruptions
    - Easy to record and observe
  - In the field
    - Realistic context
    - Difficult to carry out
    - Hard to control
Wizard of Oz

- We want our study to be as close as possible to the envisioned final application
- Chicken and egg problem: Tests of a new system that cannot be built, yet
- May need several experimental setups to test different configurations, situations, etc.

→ Wizard of Oz Experiment:
  - Users think to interact with an autonomous system
  - A human operator (wizard) controls the system
  - The users are not conscious about this fact
  - Wizard may or may not be restricted in his perception or actions to mimic the possibilities of an imperfect system
After the design phase, a new interface needs evaluation.

- Does the interface address the targeted problem?
- Does this improve the usability of the interface?

Evaluation consists of two steps:
- Evaluation of technical components
- Evaluation of interface as communicative artifact
Technical Evaluation

- Innovative user interfaces often rely on imperfect components
  - Example: Speech interface
    - Speech-/Non-Speech-Segmentation
    - Speech recognition
    - Language understanding
- Their quality has an impact on the quality of the system
- \[\Rightarrow\] We need to measure the quality of such components
  - Define a metric to compare desired result to the result of the system
    - Example: Discrete metric $\Rightarrow 0$ if correctly processed, $1$ otherwise
  - Desired result often hand-labeled by humans
    - Example: Episodes of interaction annotated for emotions
  - Measure this distance on a test data set
  - Report statistics on this performance measure
    - Example: Recognition rate (average distance on test set)
Challenges in Technical Evaluation

- Many technical components require training data
  - Testing on the training data leads to over-specialization
  - But collection of new data is very expensive
  - Is data from the lab actually representative?

- Technical components usually have many free parameters
  - Example: Number of features, number of recognized classes
  - Need to know the influence of each parameter on performance
  - There may be interaction effects between parameters
  - Not advisable to compare two mediocre systems
Interaction Evaluation

- Does the new interface improve the interaction for real users?
- Not directly connected to technical performance
- Usually done by evaluating user studies
- Two approaches:
  - Objective performance measures
    - Throughput
    - % of successfully completed tasks
  - Subjective performance measures
    - User satisfaction
    - User experience
    - Joy of use
For both technical and interaction evaluation, we are in a similar situation:

We want to compare

- a baseline system A (old reference system)
- a new system B

Technical and interaction evaluation generate sets of data

- Measure "Recognition rate" for baseline system A and new system B
- Measure "User satisfaction" for baseline system A and new system B

How do we analyze this data?

Can we draw reliable conclusions or can we only guess?

→ Test theory

→ Statistical analysis
Statistical Analysis: Example

- Example:
  - User satisfaction score from 0 to 5 for two systems A and B
  - Collected data samples from test users:
    - A: [2,3,1,1,0,2,3,...]
    - B: [3,4,3,2,1,2,4,...]

- How to analyze this data?

- First try: Calculate the arithmetic mean of the scores

- Is B better than A?
Problem: Outliers & Sample Size

- We imagine this picture...
  
  ![Diagram](image1)

- ...is caused by a distribution of data points like this:
  
  ![Diagram](image2)

- But what if the data actually looks like this:
  
  ![Diagram](image3)

- Some measurements are prone to outliers (e.g. corrupt data)
- Especially a problem with small data sets
Problem: Variance

- Assume we controlled for outliers in our data set.
- We imagine this picture:

```
0  A  B  5
```

- Is caused by a distribution of data points like this:

```
0  XXXXXXX  XXXXXXX  5
```

- But we only look at a sample of the whole population!
- What if the data looks like this:

```
0  XX  XXX  XXXXXXXXX  XX  X  5
```

- High variance within the sample
- How reliable are conclusions we draw from this data?
Problem: Independence

• Assume we controlled for outliers and variance in our data set
• Sample data is actually distributed like this:

\[ \begin{array}{c}
\text{score} \\
0 \quad \text{x} \text{x} \text{x} \text{x} \text{x} \text{x} \quad \text{x} \text{x} \text{x} \\
5
\end{array} \]

• Can we now conclude that B is better than A?
• Errors can also occur in the experiment setup!
• Maybe, we always have one participant first try A, than B
  • \( \rightarrow \) For system B, people are more experienced
  • Samples for A and B cannot be assumed to be independent
• Reversing the order could lead to a much less pronounced effect:

\[ \begin{array}{c}
\text{score} \\
0 \quad \text{x} \text{x} \text{x} \text{x} \text{x} \text{x} \quad \text{x} \text{x} \text{x} \\
5
\end{array} \]
Summary

• Iterative process of design and evaluation

• Formative evaluation
  • Identify strategies for interface improvement
  • Main tool: user studies

• Design
  • Components of an interface system
  • Statistical pattern recognition

• Summative evaluation
  • Evaluation of technical components
  • Evaluation of communicative artifact
  • Introduction of statistical analysis