

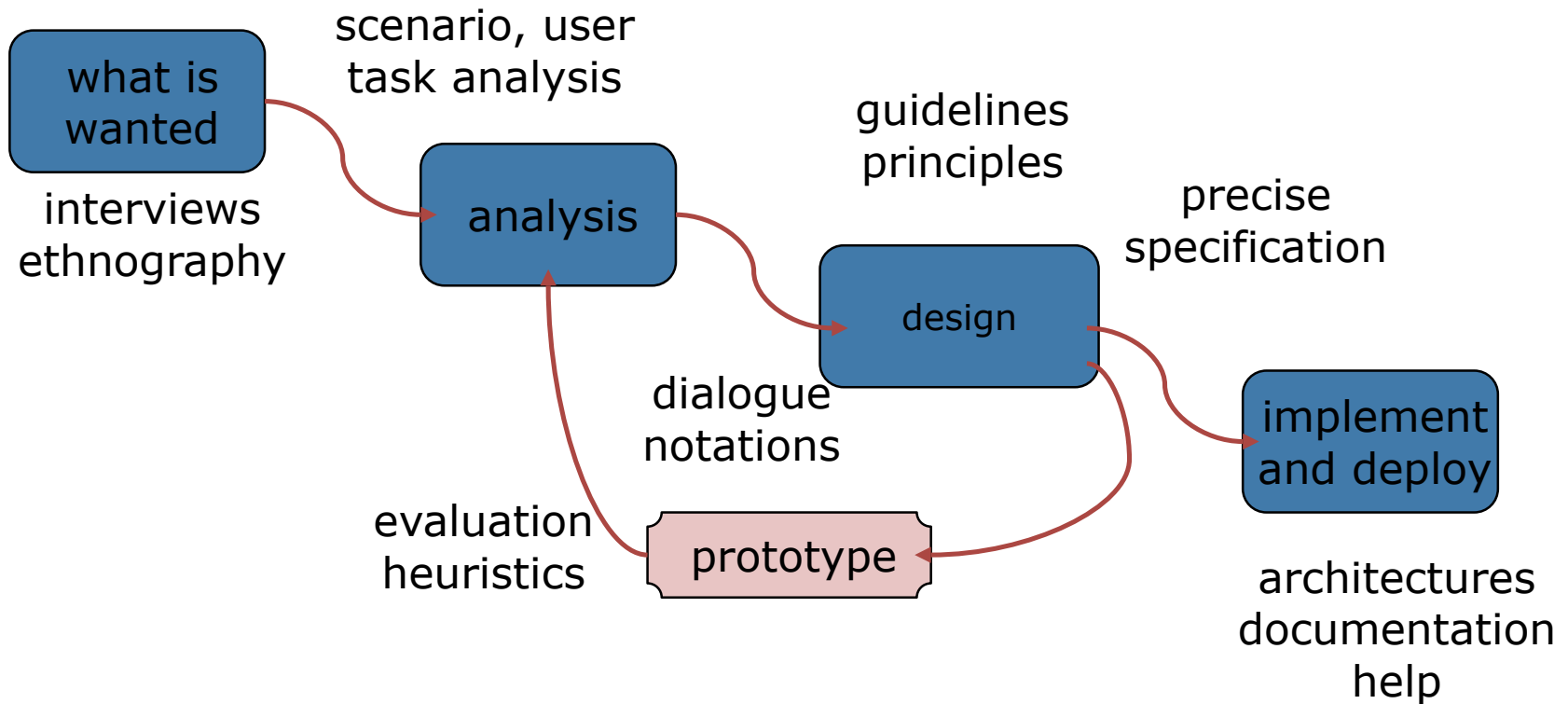
# Human-Computer Interaction

---

Termin 7:  
Usability Evaluation

# User centered design

Process to develop interactive systems such that **usability** will be maximized.



# Prototyping

- Je früher Prototyp bauen und **testen**, desto besser
- Horizontale vs. vertikale Prototype:
  - horizontal: vollständige Schnittstelle, keine Funktionalität
  - vertikal: teilweise Funktionalität realisiert
  - Mischformen sinnvoll und üblich
- Stufen des Prototyping
  - **Konzeptueller Prototyp**: User erhält Beschreibung und stellt sich Funktionsweise des Systems vor
  - **Papierprototyp**: Skizzen, Bilder, etc.
  - **Statische Bildschirmdarstellungen**: einzelne Darstellungen
  - **Dynamische Simulation**: Simulierte einfache Abläufe
  - **Wizard-of-Oz**: Simulation durch unsichtbare Person



# Key questions for today

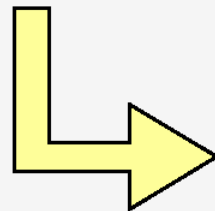
- How can the usability of a system be evaluated?
- How can usability problems be found?



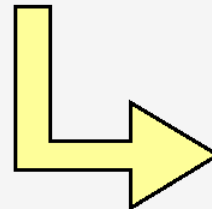
# Evaluieren: Warum und was?

## 6.3.1 Systematik

**Ziel**  
Wozu soll evaluiert werden?



**Kriterium**  
Was soll evaluiert werden?



**Mittel**  
Wie soll evaluiert werden?

Evaluation  
ist die Überprüfung eines  
konkreten Systems auf  
Übereinstimmung mit  
festgelegten Kriterien



# Evaluation

- **Warum?** Bewerte Usability und Effekte auf Nutzer, finde Probleme, gib Hinweise für Verbesserungen
- **Was?** Usability-Kriterien
- **Wo?** Labor oder Umgebung/Feldstudie
- **Wer?** Experte (mit/ohne Nutzerbeteiligung) oder echte Nutzer
- **Wann?** In *allen* Stadien des Entwicklungszyklus (Ideen, Prototypen, System)
  - **Formative Evaluation:** verschiedene Zeitpunkte, aktuelles System gegenüber den Anforderungen
  - **Summative Evaluation:** abschließende Bewertung hinsichtlich vorher formulierter Kriterien



# Evaluation - Grundsätzlicher Ablauf

- Festlegung von **Kriterien**
  - die das System erfüllen muss, um als „usable“ zu gelten
- Festlegung von **Leistungsstandards**
  - die das System erreichen muss, um ein Kriterium zu erfüllen (Operationalisierung)
- **Messung** und Vergleich (Analyse)
  - Untersuchung unter Anwendung der Kriterien und Vergleich auf vorgegebene Leistungsstandards
- **Werteurteil** (Synthese)
  - Verknüpfung von Ergebnissen zu einem Urteil
  - Vorschläge zur Verbesserung



# Methodenwahl und Evaluationsdesign

- Gültigkeit** (Validität)
  - wird das gemessen/beobachtet, was man messen wollte ?
- Zuverlässigkeit** (Reliabilität)
  - Ist die Studie wiederholbar?
- Signifikanz** und **Verallgemeinerbarkeit**
  - Auswahl der Teilnehmer, beeinflusst Kontext der Studie das Verhalten???

## Pilot/Vorstudien

- Im Ernstfall immer zunächst eine Pilotstudie durchführen
- Durchführbarkeit testen, Verlauf & Moderation üben, verbessern
- Notfalls Kollegen als Versuchspersonen nehmen
- Eventuell sogar mehrere iterative Vorstudien





# Evaluationsmethoden

- Häufig gehörte Begriffe/Unterscheidungen:
  - harte vs. weiche
  - formative vs. summative
  - qualitative vs. quantitative
  - subjektive vs. objektive
  - analytische vs. heuristische vs. empirische
  - formale vs. informelle
  - experimentelle vs. leitfadenorientierte
  - theory-based vs. user-based



# Evaluationsmethoden

## □ Inspektionsmethoden (Experten-Review)

- Guidelines review & consistency inspection
- Cognitive walkthrough
- Heuristic evaluation
- Focus group

## □ Benutzerstudien

- Usability-Test
- Thinking-Aloud
- Field studies
- Interviews & questionnaires

## □ Model-basierte Evaluation



# Usability-Inspektionsmethoden

---

Guidelines Review  
Consistency Inspection  
Cognitive Walkthrough  
Heuristic Evaluation  
Focus Group

# Guideline review

## Consistency inspection

- System/interface is checked for conformance with guidelines
  - Standard guidelines, e.g. Shneiderman's rules
  - Organization specific guidelines, e.g. Apple styleguide
  
- Consistency inspection
  - of terminology, colors, fonts, icons, menus, general layouts, etc.
  - of interaction style



# Cognitive Walkthrough

- Aufgabenorientierte Inspektionsmethode („Benutzbarkeits-Gedankenexperiment“)
- Usability-Experte untersucht Funktionalität im Interesse eines imaginären Benutzers
  - wählt Aufgabe, die das System unterstützen soll
  - „spielt“ Aufgabe durch (*walks through*)
  - bestimmt korrekte Handlungsabläufe und identifiziert potentielle Probleme eines Benutzers
- Vorteile:
  - früh einsetzbar, Fehler im Ansatz erkennbar
- Nachteile:
  - Kann der Evaluator den Benutzer einschätzen?



# Cognitive Walkthrough - Durchführung

## Vorbereitungsphase

- Detaillierte Beschreibung der potentiellen Nutzer (Wer, Vorwissen, ...)
- Genaue Aufgabenstellungen unterteilt in die nötigen Einzelschritte
- Liste der möglichen Aktionen und der Auswirkungen
- Angewandt auf Prototyp oder Beschreibung des Systems

## Analyse

- Experten gehen durch alle Aktionen und Systemantworten und beantworten jeweils die Fragen:
  - Sind die richtigen Aktionen verfügbar (Effekte = Nutzerziele )?
  - Werden Nutzer merken, dass die richtigen Aktionen verfügbar sind?
  - Werden sie die korrekten Aktionen erkennen?
  - Werden die Nutzer das System-Feedback verstehen?

## Follow-Up

- Aufzeichnung von Ergebnissen und Überlegungen zu Designalternativen und Verbesserungsvorschläge



# Beispiel: Ergebnisse eines Experten-Reviews des Otto-Versands

The screenshot shows a search results page on the Otto-Katalog website. The browser window is titled 'Otto-Katalog - Microsoft Internet Explorer'. The address bar shows 'http://www.otto.de/otto/cgi-bin/ottohome?MainWin=artikelein'. The search results are displayed under the heading 'Ihr Suchergebnis'. There are three search results, each with a small image and a title. The first result is 'Organizer, Zub., Zubehör'. The second is 'Organizer, Zubehör, Zub. > Pilote-Modelle, Zub.'. The third is 'Packard Bell Desktop, Packard Bell Monitor'. There are navigation buttons for 'vorherige Seite' and 'nächste Seite'. At the bottom of the page, there are several navigation links: 'Unser Sortiment', 'Warenkorb/Bestellen', 'Suchen', 'Bestell-Nr. eingeben', 'Information', and 'Homepage'. The page is annotated with several yellow boxes containing text, with lines pointing to specific elements on the page.

Annotations:

- Artikelübersicht nicht erkennbar
- Gesuchte Begriffe nicht angezeigt
- Nicht abgegraut
- Navigations-elemente an ungewohnter Stelle
- Bilder sehr klein
- Umfang des Suchergebnisses fehlt
- Sortierung nicht möglich
- Unverständliche Abkürzungen
- Zu dicht beieinander



# Verbesserungsvorschläge des Experten-Reviews

Herbst/Winter '97/98

### Bestell-Nr. eingeben

Bitte geben Sie Anzahl, Artikelnummer und die gewünschten Lieferservices in die Tabelle ein!

Anzahl:	<input type="text"/>	Artikelnummer:	<input type="text"/>
Größe:	<input type="text"/>	Andere Lieferanschrift:	<input type="text"/>
Zahlung:	<input type="text" value="in 1 Rate"/>	Lieferservice:	<input type="text"/>
		Wunschtermin:	<input type="text"/>

In den Warenkorb übernehmen und weitere Artikel  
Nur diesen Artikel in den Warenkorb übernehmen

Unser Sortiment | Warenkorb/Bestellen | Suchen | Bestell-Nr. eingeben | Informationen

Sie suchen etwas Bestimmtes? Unsere Suche bietet Ihnen umfangreiche Möglichkeiten, es auch zu finden

Original

Empfehlung

Herbst/Winter '97/98

### Bestell-Nr. eingeben

Bitte füllen Sie das Formular aus.

Artikel:	Bestellnummer:	<input type="text"/>
	Menge:	<input type="text"/>
	Größe:	<input type="text"/>
Lieferung:	Lieferservice:	<input type="text" value="Normale Lieferung"/>
	Wunschtermin:	<input type="text"/>
	Andere Lieferanschrift?	<input type="checkbox"/>
Zahlung:	Raten:	<input type="text" value="eine"/>

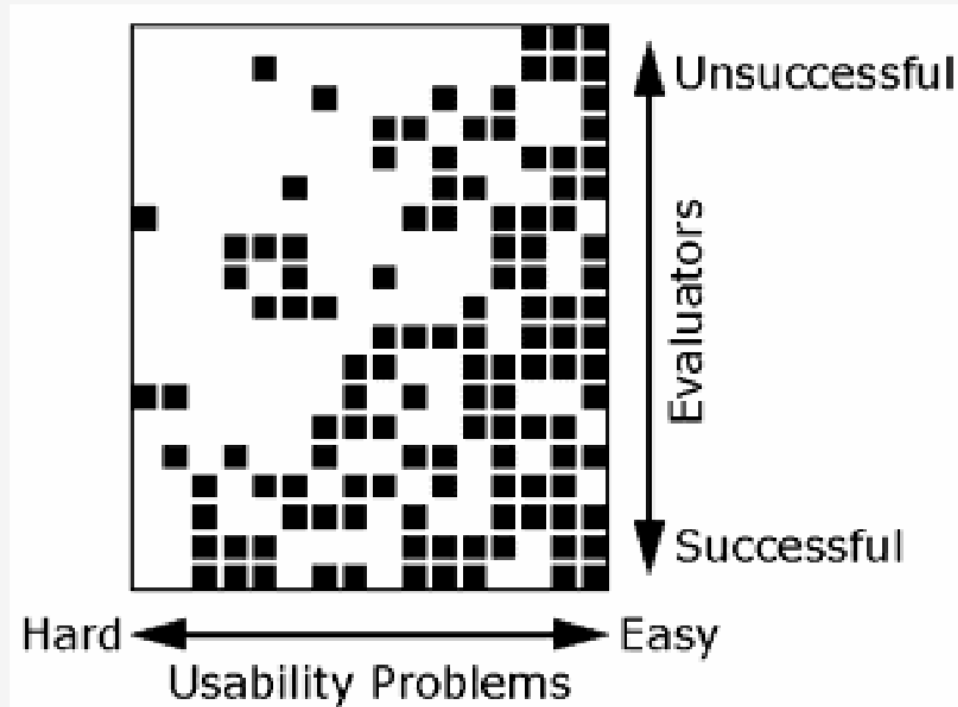
OK | Eingabe verwerfen

Unser Sortiment | Warenkorb/Bestellen | Suchen | Bestell-Nr. eingeben | Informationen





## Wieviele Reviewer ?



19 Reviewer sollten  
16 Fehler in einer  
Bankensoftware finden:

- Unterschiedliche Reviewer fanden durchaus unterschiedliche Fehler
- Die erfolgreichsten Reviewer finden nicht immer die schwierigsten Fehler

## Einsatz mehrerer Reviewer sinnvoll !

Nielsen, J.: How to conduct a heuristic evaluation, [http://www.useit.com/papers/heuristic/heuristic\\_evaluation.html](http://www.useit.com/papers/heuristic/heuristic_evaluation.html)

Optimal: 4 Reviewer - Nutzen 62 mal größer als Kosten

5 Reviewer erkennen 75-80 % Fehler – gut, aber:

-> nicht im Kernkraftwerk anwenden!



# Heuristic Evaluation

- Usability-Experten bewerten System anhand einfacher und allgemeiner Usability-Heuristiken
- Unabhängig von mehreren Experten durchgeführt; Daumenregel: 5 Experten finden 75% der Probleme
- Testen entweder lauffähiges System oder Prototypen
  
- Heuristiken/Kriterien:
  - Nielsen's 10 Heuristiken (1993; siehe Vorlesung 6)
  - Erweiterte Heuristiken ab 2001 (Nielsen, 2001)



# Usability-Heuristiken (1)

- **Visibility of system status**
  - System should always keep users informed about what is going on
- **Match between system and the real world**
  - Should speak the users' language, with words, phrases and concepts familiar to the user. Follow real-world conventions, making information appear in a natural and logical order.
- **User control and freedom**
  - Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state (Support undo and redo)
- **Consistency and standards**
  - Users should not have to wonder whether different words, situations, or actions mean the same thing.



# Usability-Heuristiken (2)

## □ **Error prevention**

- Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

## □ **Recognition rather than recall**

- Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another.

## □ **Flexibility and efficiency of use**

- Accelerators --unseen by the novice user --often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.



# Usability-Heuristiken (3)

- **Aesthetic and minimalist design**
  - Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
- **Help users recognize, diagnose, and recover from errors**
  - Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
- **Help and documentation**
  - Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.



# Heuristic Evaluation - Durchführung

- Trainingssitzung
  - Gutachter üben (falls nötig) detailliert Heuristiken
  
- Evaluation
  - Jeder Evaluator begutachtet mit Liste von standardisierten Heuristiken mehrmals das Interface - in der Regel 4 Durchgänge
  - Nimmt Einblick in Informationsablauf und Funktionalität einzelner Bedienelemente (*nicht* aufgaben-orientiert)
  - Beobachter protokolliert erkannte Probleme
  - Kommunikation der Evaluatoren erst nach Durchgängen



# Heuristic Evaluation - Durchführung

- Ergebnisse und Abschlusssetzung
  - Liste der Probleme (verletzte Prinzipien + Begründung)
  - möglichst detaillierte Beschreibung der Probleme
  
- Problembewertung
  - Wie ernst ist das Usability-Problem? Ist es zu umgehen?
  - Jeder Gutachter bekommt Liste von Problemen und gibt Bewertung der Ernsthaftigkeit ab:
    - 0 - don't agree that this is a usability problem
    - 1 - cosmetic problem
    - 2 - minor usability problem
    - 3 - major usability problem - important to fix
    - 4 - usability catastrophe; imperative to fix
  - Abschließendes Ranking der Probleme



# Heuristic Evaluation

## □ Beispiel:

- *Interface used command „Save“ on 1st screen for saving the user's file, but used „write file“ on 2nd screen. Users may be confused by this different terminology.*
- Violation of consistency/standards - severity rating 3

## □ Vorteile:

- Schnell, kostengünstig, qualitativ gute Ergebnisse

## □ Nachteile:

- Gutachter sind nicht Benutzer
- Heuristiken erfassen nicht alle Probleme, die später beim Benutzer vermutlich auftauchen





# Focus Group

- Moderierte Gruppendiskussion mit ausgewählten Endnutzern
- Gruppe sollte homogen sein für guten Kommunikations- und Meinungsbildungsprozess
- Ziel: Sammeln von Kundenmeinungen, Anregungen für weitere Entwicklungen



# Benutzerstudien/User studies

---

Thinking-Aloud  
Cooperative Evaluation  
Interviews & questionnaires  
Usability-Test

# User studies

- ❑ Study the interaction between **actual user** and system
- ❑ Involves measuring **representative users' performance** on **carefully prepared tasks** that are typical of the tasks, for which the system was designed
- ❑ Users testing can use
  - video and interaction logging to capture errors and frequencies and time of commands
  - think-aloud protocols
- ❑ User testing may be done in the lab or the field
- ❑ Users may be interviewed or complete questionnaires
  - Satisfaction questionnaires provide data about users' opinions



## Lab studies

- Experiment under controlled conditions
  - specialist equipment available
  - uninterrupted environment
- Disadvantages:
  - lack of context
  - difficult to observe user cooperation
- Prevalent paradigm in psychology

## Field studies

- Experiments dominated by group formation
- Field studies more realistic
  - *distributed cognition* ⇒ work studied in context
  - real action is *situated*
  - physical *and* social environment crucial
- sociology and anthropology – open study and rich data



# Think Aloud

- User is observed while performing a *predefined* task
- User is asked to describe what...
  - he is thinking right now
  - he is expecting to happen
  - he is thinking is happening
  
- Advantages
  - simplicity - requires little expertise
  - can provide useful insight into user's mental model
  - can show how system is actually used
- Disadvantages
  - artificial test situation → cooperative evaluation
  - subjective and selective → several trials needed
  - act of describing may alter task performance



# Cooperative Evaluation

- User evaluates together with expert, sees himself as collaborator in evaluation
- both can ask each other questions
  
- Additional advantages
  - less constrained and easier to use
  - user is encouraged to criticize system
  - clarification dialogues possible
  
- Problems with *both* techniques
  - generate a large volume of information (*protocols*)
  - 'Protocol analysis' crucial and time-consuming



# Query techniques

## □ Interviews:

- analyst questions user, usually based on prepared questions
- informal, subjective and relatively cheap
- can be varied to suit context, issues can be explored more fully, can unanticipated problems
- very subjective, time consuming

## □ Questionnaires:

- Set of fixed questions given to users, need careful design!
- Style of questions: open vs. closed, scalar (judge a statement on a numeric scale), multiple-choice, ordering, negative vs. positive, ...
- Style of answers: text, ja/nein, number of options, ...
- reaches large user group, can be analyzed rigorously, less flexible, less probing



# Surveys

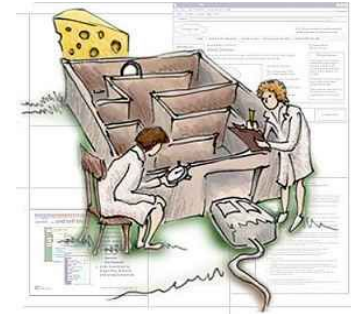
- Written user surveys (usually online)
  
- Useful to gather important data about
  - user background (age, gender, education, etc.)
  - experiences with computers
  - job responsibilities
  
- Can use survey instruments to determine
  - personality style of user
  - familiarity with features
  - feeling state of using the system





# Usability Testing

- Aufnehmen typischen Benutzerverhaltens bei typischen Aufgaben in kontrolliertem Szenario. Auch Feldbeobachtung
- Benutzer werden bei Aufgabenbearbeitung beobachtet und auf Video aufgenommen, Tasten/Mausbewegung “geloggt”
- Daten genutzt, um Bearbeitungszeit zu berechnen, häufige Fehler zu entdecken, erkennen, warum User etwas tun
- “User Satisfaction” Fragebögen und Interviews für subjektive Meinungsäußerung.



# Usability Test - Durchführung

- Suche repräsentative Benutzer
  - 5-10 Benutzer als Testpersonen (abhängig von Budget, Zeitplan, Verfügbarkeit Testpersonen)
- Kriterien der Auswertung auswählen (Beispiele):
  - Zeit für Aufgabenerfüllung
  - Zeit für Aufgabe nach Ablenkung/neuem Input
  - Anzahl und Art von Fehlern pro Aufgabe oder pro Zeiteinheit
  - Anzahl Zuhilfenahme Onlinehilfe oder Manual
  - ...
- Entwickle Testszenarien
  - relevante Szenarien (typische Aufgaben vs. Extremsituationen)
  - Aufgaben kürzer als 30 Minuten
  - Identische Testbedingungen für alle
- Ethische Fragen
  - Probanden Aufklären, Einverständniserklärung, etc.



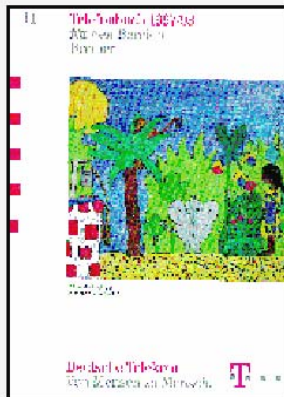
# Usability Test - Durchführung

- Vorab: Pilottests
  - Schulung von Experimentatoren und Beobachtern
- Eigentlicher Test
  - Einführung/Erläuterung des Tests für die Versuchspersonen
  - Testdurchführung und Datenaufzeichnung
  
- Auswertung
  - Statistiken, z.B. Maus-Events, Menü-Auswahlen
  - Bildschirm-Layout: Blickverfolgung und Aufgabenablauf
  - Post-task Videokonfrontation und User-Interview
  
- Vermittlung der Ergebnisse an Entwickler



# Usability Test - Beispiel

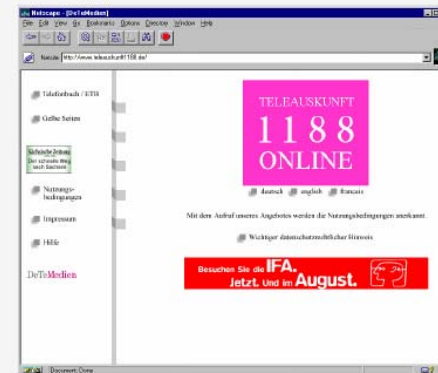
- Ziel: Vergleich unterschiedlicher Telefonauskunftssysteme
  - hinsichtlich ihrer Benutzbarkeit.
  - Verfahren: Vier Versuchspersonen bearbeiten jeweils 4 Prüfaufgaben.
  - Die Bearbeitung wird mit Video, Audio und Logging-Programmen protokolliert.



Telefonbuch



Telefon-CD  
der DeTeMedien



www.teleauskunft1188.de

# Ergebnisse Beobachterkommentare

The screenshot shows a Netscape browser window displaying a 'Telefonbuch' (phone book) page. The page has a left sidebar with navigation links like 'Telefonbuch / ETB', 'Gelbe Seiten', 'MediaGlobe', 'Impressum', and 'Hilfe'. The main content area contains a form for entering contact information (Name, Ort, PLZ, Vorname, Vorwahl, Straße) and search options (Suchformen: Standard, Nahbereich, Ähnliche Namen, etc.). There are also buttons for 'Suchen' and 'Löschen'. A red 'SPIEGEL ONLINE' logo is visible on the right. The browser's address bar shows the URL 'http://www01.teleauskunft1188.de/cgi-bin/tron.cgi'.

Annotations in yellow boxes:

- Nicht clickbare Knöpfe hervorgehoben: **regelmäßige Fehlversuche**
- Kleine Knöpfe für häufig genutzte Funktionen: **häufiges Zögern**
- Ungewöhnliche Feldreihenfolge: **häufige Fehleingaben**
- Suchformen werden von keinem Probanden verstanden: **Nicht genutzt**
- Großer, prominent positionierter Knopf: **Nur einmal gedrückt**



# Zeitdauer & Korrektheit im Vergleich

## Zusammengefaßte Ergebnisse



Aufgabenstellung				
1. Suche die Telefonnummer von Maria Müller. Sie wohnt Am Ziegelberg in Bremen.	Korrekte Ergebnisse	★★★★	★★★	★★★★★★
	Bearbeitungsdauer [min]	0:45	2:30	3:00
2. Suche die private Telefonnummer von Carsten Bormann (TZI-Bereich Digitale Medien und Netze).	Korrekte Ergebnisse	★	★	
	Bearbeitungsdauer [min]	0:30	1:00	2:45
3. Marc-Oliver Schulze wohnt bei seinem Vater in Bremen. Seine Telefonnummer beginnt mit einer "40".	Korrekte Ergebnisse	★★★★★★	★★★★★	
	Bearbeitungsdauer [min]	1:15	1:50	4:10
4. Suche einen Sportarzt in Bremen.	Korrekte Ergebnisse	★★★★	★	★★
	Bearbeitungsdauer [min]	0:30	2:30	4:20

Beobachtung  
Usability Test

# Physiological measurements

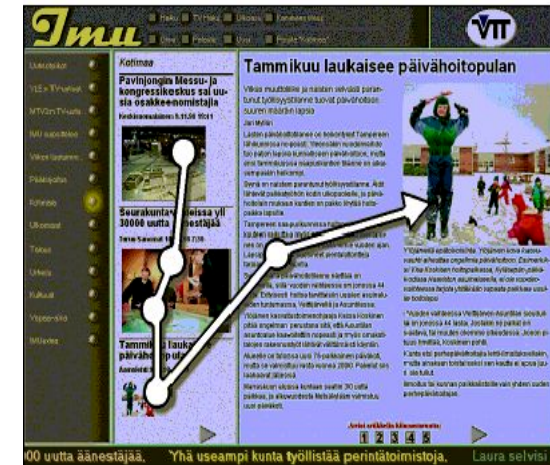
- Emotional response linked to physical changes
- may help determine a user's reaction to an interface
- measurements include:
  - heart activity, including blood pressure, volume and pulse.
  - activity of sweat glands: Galvanic Skin Response (GSR)
  - electrical activity in muscle: electromyogram (EMG)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





# Eye tracking

- head or desk mounted equipment tracks the position of the eye
- eye movement reflects 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





# Recall: Methods in user-centered design

- Field studies (including contextual inquiry)
- User requirement analysis
- Iterative design
- Usability evaluation
- Task analysis
- Focus groups
- Formal heuristic evaluation
- User interviews
- Surveys
- ...

Ranking based on survey among experienced UCD practitioners  
(103 questionnaires) (Mao et al., 2005)



# Summary - Choosing an Evaluation Method

when in process:	design vs. 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

