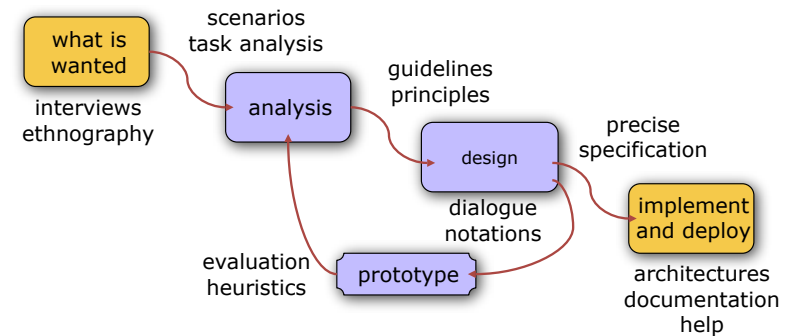


Human-Computer Interaction

Session 7 Usability Evaluation

The design process

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



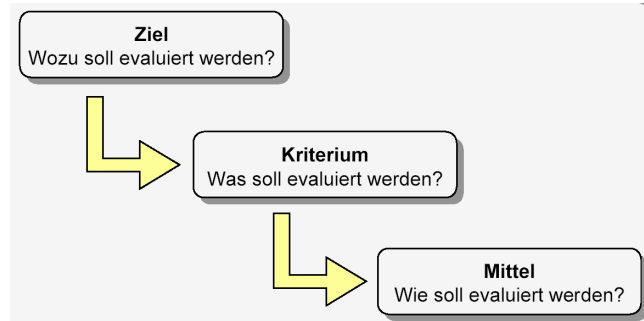
Prototyping

- Je früher Prototyp bauen und testen, desto besser
- **Horizontale vs. vertikale Prototypen:**
 - horizontal: vollständige Schnittstellen, 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**?

Evaluation = Überprüfung eines konkreten Systems auf Übereinstimmung mit vorher 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

1. Festlegung von **Kriterien**
 - die das System erfüllen muss, um als „usable“ zu gelten
2. Festlegung von **Leistungsstandards**
 - die das System erreichen muss, um ein Kriterium zu erfüllen (Operationalisierung)
3. **Messung** und Vergleich (Analyse)
 - Untersuchung unter Anwendung der Kriterien und Vergleich auf vorgegebene Leistungsstandards
4. **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

- Usability inspection (**expert review**)
 - Guidelines review & consistency inspection
 - Cognitive walkthrough
 - Heuristic evaluation
 - Focus group
- **User studies**
 - Usability-Test
 - Thinking-Aloud
 - Field studies
 - Interviews & questionnaires
- **Model-based evaluation**

Usability inspection methods

Guidelines Review
Consistency Inspection
Cognitive Walkthrough
Heuristic Evaluation

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

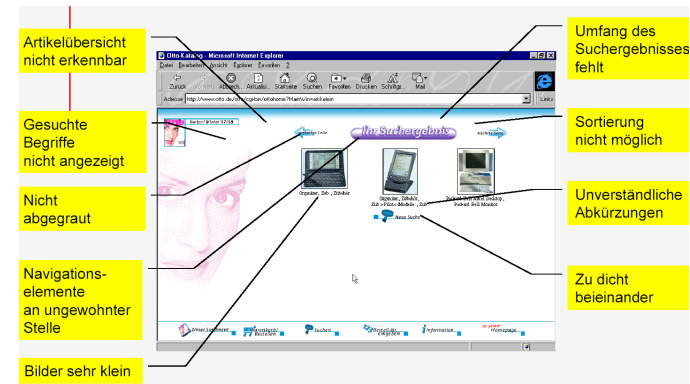
Analyse

- 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?

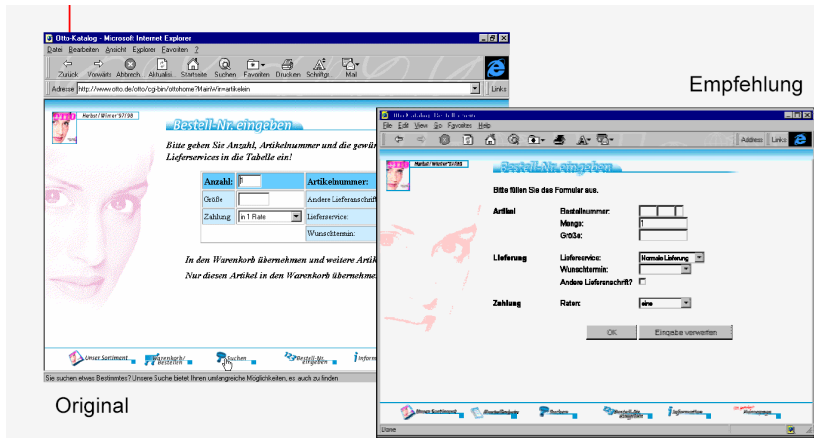
Follow-Up

- Aufzeichnung von Ergebnissen und Überlegungen zu Designalternativen und Verbesserungsvorschläge

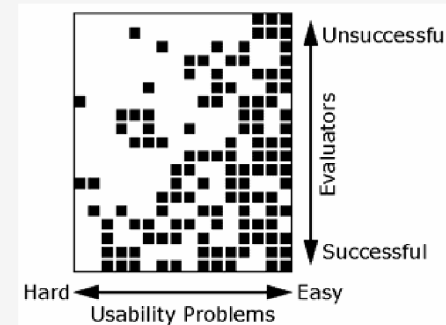
Beispiel: Ergebnisse eines Experten-Reviews des Otto-Versands



Verbesserungsvorschläge des Experten-Reviews



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

Optimal: 4 Reviewer - Nutzen 62 mal größer als Kosten
 5 Reviewer erkennen 75-80 % Fehler – gut, aber:
 -> nicht im Kernkraftwerk anwenden!

Heuristic Evaluation

J. Nielsen (1993)
www.useit.com

- 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**
- **Match between system and the real world**
 - Speak the users' language, follow real-world conventions, make information appear in a natural and logical order
- **User control and freedom**
 - Provide a clearly marked "emergency exit" to leave an unwanted state (undo and redo)
- **Consistency and standards**
 - Users should not have to wonder whether different words, situations, or actions mean the same thing.
- **Error prevention**

Usability-Heuristiken (2)

- **Recognition rather than recall**
- **Flexibility and efficiency of use**
 - cater both inexperienced and experienced users, allow to tailor frequent actions
- **Aesthetic and minimalist design**
 - provide no irrelevant or rarely needed info
- **Help users recognize, diagnose, and recover from errors**
 - Error messages in plain language (no codes), precisely indicate the problem, suggest a solution.
- **Help and documentation**
 - provide help and documentation, easy to search, focus on user task, list concrete steps to be carried out, not 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 Abschlussitzung
 - 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

Benutzerstudien/User studies

Thinking-Aloud
Cooperative Evaluation
Interviews & questionnaires
Usability-Test

User studies

- Study 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
- Can use video and interaction logging to capture errors and frequencies and time of commands, or think-aloud protocols
- May be done in the lab or the field
- Users may be interviewed or complete questionnaires
 - gather 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 exp. 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 → multiple trials & users needed
 - act of describing may alter task performance

Cooperative Evaluation

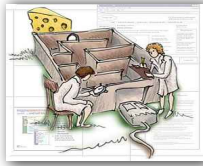
- User evaluates together with expert, sees himself as collaborator in evaluation
 - both can ask each other questions
- Additional advantages
 - less constrained and easier to use
 - user is encouraged to criticize system
 - clarification dialogues possible
- Problems with *both* techniques
 - generate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

Query techniques

- **Interviews:**
 - analyst questions user, based on prepared questions
 - informal, subjective, and relatively cheap
 - can be varied to suit context, issues can be explored more fully, can reveal unanticipated problems
 - very subjective and 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, yes/no, number of options, ...
 - reaches large user group, can be analyzed rigorously, less flexible, less probing

Usability Testing

- Aufnehmen typischen Benutzerverhaltens bei typischen Aufgaben in kontrolliertem Szenario
- 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
- "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 vs. Extremsituationen)
 - Halte 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.



Ergebnisse Beobachterkommentare

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 | Korrekte Ergebnisse | ★★★★ | ★★ | ★★★★★ |
|--|--------------------------|-------|------|-------|
| 1. Suche die Telefonnummer von Maria Müller. Sie wohnt Am Ziegeberg in Bremen. | Korrekte Ergebnisse | ★★★★ | ★★ | ★★★★★ |
| | Bearbeitungs-dauer [min] | 0:45 | 2:30 | 3:00 |
| 2. Suche die private Telefonnummer von Carsten Bormann (TZ-Bereich Digitale Medien und Netze). | Korrekte Ergebnisse | ★ | ★ | |
| | Bearbeitungs-dauer [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 | ★★★★★ | ★★★★ | |
| | Bearbeitungs-dauer [min] | 1:15 | 1:50 | 4:10 |
| 4. Suche einen Sportart in Bremen. | Korrekte Ergebnisse | ★★★★ | ★ | ★★ |
| | Bearbeitungs-dauer [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 and pulse
 - activity of sweat glands: Galvanic Skin Response (GSR)
 - electrical activity in muscle: electromyogram (EMG)
 - electrical activity in brain: electroencephalogram (EEG)
- often difficult to interpret these physiological responses

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



Remember, methods in UCD

1. **Field studies**
2. User requirement analysis
3. Iterative design
4. **Usability evaluation**
5. Task analysis
6. **Focus groups**
7. **Formal heuristic evaluation**
8. **User interviews**
9. **Surveys**
10. ...

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

Statistical analysis

Messen & Werte - Skalen

- *Skalen* für Messwerte in Typenhierarchie einteilbar
- Typ einer Skala wird auch als *Skalenniveau* bezeichnet
- man unterscheidet die Niveaus
 1. *Nominalskala*: gleiche Merkmalsausprägung, gleiche Werte: Blutgruppen A, B, AB, Null; Geschlecht m., w.
 2. *Ordinalskala*: größerer Merkmalsausprägung, größere Werte: Kleidungsgrößen S, M, L, XL, XXL; Urteile --, -, 0, +, ++
 3. *Intervallskala*: Größenordnung von Wertdifferenzen entspricht Merkmalsunterschieden: Temperatur in °C oder Fahrenheit
 4. *Verhältnisskala*: Größenverhältnisse der Werte entspricht Merkmalsausprägungen: Länge in mm
 5. *Absolutskala*: absolute Werte entsprechen Merkmalsausprägungen: Anzahl der Kinder

Statistische Evaluation

- **Deskriptive Statistik** (beschreibend) für quantitative Erhebungen, etwa
 - Daten von Benutzerbefragungen
 - Daten von Messungen (z.B. Zeit zur Bearbeitung von Aufgaben)
 - Veranschaulichung von Zahlenmaterial
- **Inferenzstatistik** (Test von Hypothesen), etwa
 - konkurrierende Schnittstellen-Designs (Bsp.: "GUIs mit Zeichnungen als Icons sind einfacher zu bedienen als GUIs mit Fotos")
 - verschiedene Benutzergruppen (Bsp.: "ältere Benutzer haben mehr Probleme mit der Lesbarkeit als jüngere")
 - verschiedene Eingabemodalitäten (Bsp.: „Multimodale Eingabe mit Sprache und Gestik ist effizienter als direkte Manipulation per Spacemouse“)

Deskriptive Statistik

Zentrale Frage der deskriptiven Statistik:

Wie können die empirischen Daten aufbereitet, dargestellt, zusammengefasst und strukturiert werden, so dass zentrale Merkmale sichtbar werden?

Darstellung von Häufigkeitsverteilungen durch Kennwerte

Häufigkeit: absolute, relative, kumulierte

Mittelwert, bei (mind.) Intervalldaten und symmetrischen Verteilungen

$$\bar{x} = \frac{\sum x_i}{n}$$

Median/Zentralwert: halbiert nach Größe geordnete Messreihe

$$\tilde{x}$$

Modus/Modalwert: am häufigsten vorkommender Wert

$$\ddot{x}$$

Zentrale Tendenz am Beispiel

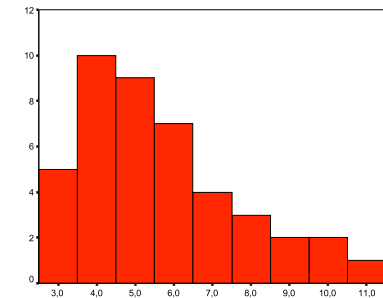
Häufigkeitsverteilung für eine Variable:

| | | | | | | | | | |
|------------|---|----|---|---|---|---|---|----|----|
| Wert | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| Häufigkeit | 5 | 10 | 9 | 7 | 4 | 3 | 2 | 2 | 1 |

arithm. Mittel 5,65

Median 5

Modus 4



Kennwerte der Dispersion (Streuung)

Varianz und Standardabweichung, geeignet bei (mind.) Intervalldaten, großen Stichproben und symmetrischer Verteilung

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Quartilabstand: Differenz der Mediane der am Median geteilten Messreihe (0.25- bis 0.75-Quantil)

$$q = (|q_{0.75} - q_{0.25}|)$$

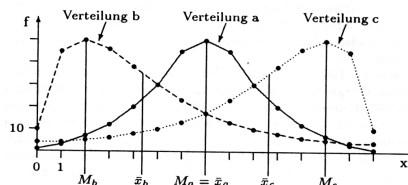
Schiefe: bezeichnet, wie stark rechts- oder linkslastig eine Verteilung ist

Pearson's Schiefeß

Bowley-Fisher's Schiefeß

$$S_p = \frac{\bar{x} - M}{s} \leftarrow \text{Modus}$$

$$S_b = \frac{\sum (x_i - \bar{x})^3 / n}{s^3}$$



$S < 0$: Verteilung ist linksschief (rechtssteil)
 $S > 0$: Verteilung ist rechtsschief (linkssteil)
 $S = 0$: Verteilung ist symmetrisch

Inferenzstatistik

Zentrale Frage der Inferenzstatistik:

Wie können auf Basis empirischer Daten (Wahrscheinlichkeits-)Aussagen über die Allgemeinheit getroffen werden?

Prinzipieller Ablauf

- Formuliere Hypothese über Einfluss unabhängiger Variablen IV auf abhängige Variablen DV
 - Arbeitshypothese: Änderung in DV verursacht durch IV
 - Null-Hypothese: Änderung in DV zufällig, kein Einfluss von IV
- Ziel: Widerlege Null-Hypothese mit minimaler Wahrscheinlichkeit eines Fehlers (H_0 liegt doch vor)
 - Irrtumswahrscheinlichkeit ergibt Signifikanzniveau: 0.05 (signifikant), 0.01 (hoch sign.), 0.001 (höchstsign.)
- Standard-Testverfahren, abhängig von Design

Model-based evaluation

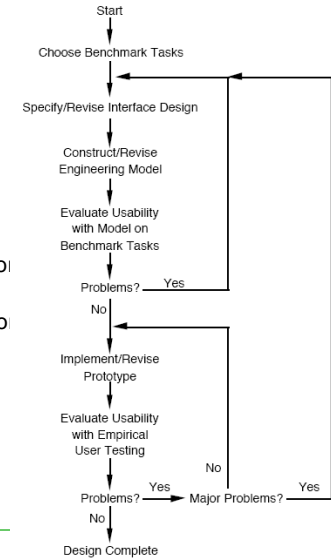
Model-based evaluation

Changes design process:

Four steps:

1. **Describe** interface design in detail
2. **Build** model of user doing a task
3. Use the model to **predict** execution or learning time
4. **Revise** or choose design depending on prediction

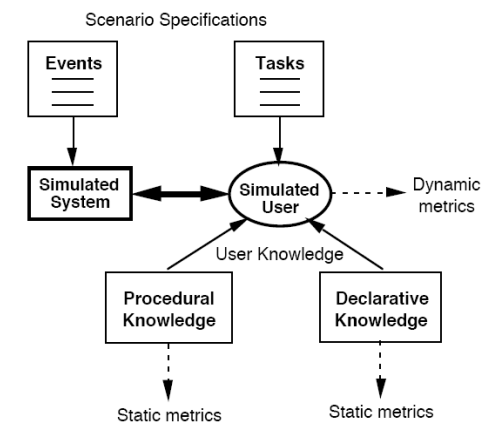
- Provides usability results *before* prototype or user testing
- Engineering the model allows more design iterations



Model-based approach

- Model summarizes interface design from the user point of view
 - Represents how *the user* gets things done with the system.
 - Components of model can be reused to represent design of related interfaces.
- *But*, current models can only predict a few aspects:
 - Time required to execute specific tasks.
 - Ease of learning of procedures, consistency effects
- Actual user testing is still indispensable!

Overview



Models = simulations of human-computer interaction

Procedural knowledge: how-to procedures → executable

Declarative knowledge: facts, beliefs → reportable

Reproducing human constraints

- If a model can be programmed to do any task at any speed or accuracy, something's wrong
- Many HCI tasks dominated by *perceptual-motor activity*
 - A steady flow of physical interaction between human and computer („doing rather than thinking“)
 - Time required depends on human characteristics and computer's behavior (determined by the design)
- Implications:
 - Modeling perceptual-motor aspects is often practical, useful, and relatively easy.
 - Modeling purely cognitive aspects of complex tasks is often difficult, open-ended, and requires research resources.

MMI / SS06

Modeling approaches

Three current approaches:

1. Task network models – before detailed design
2. Cognitive Architecture Models – packaged constraints
3. GOMS models – relatively simple & effective

Differ in

- what human constraints are modeled (cognitive/psychological vs. perceptual vs. motoric)
- level of detail
- when to use it in the design process

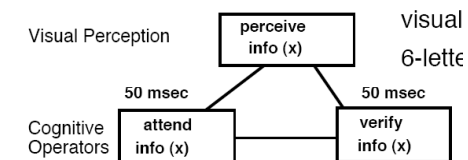
MMI / SS06

Task Network Models

- Connected network of tasks:
 - Connection: one task is a prerequisite of the other
 - Both serial and parallel execution of tasks
 - Final completion time computed from chain of serial and parallel tasks
 - *Critical path* = chain with largest execution time
- Tasks = mixture of human and machine tasks
- Each task characterized by a distribution of completion times, and arbitrary dependencies and effects

MMI / SS06

Task network - simple example



MMI / SS06

Cognitive architectures

"Programmed" with a **strategy** to perform specific tasks

- provides constraints on form and content of the strategy
- architecture + specific strategy = model of a specific task

To model a specific task

- Do task analysis to arrive at human's strategy for the task
- "Program" the architecture with representation of strategy
- Run the model using task scenarios

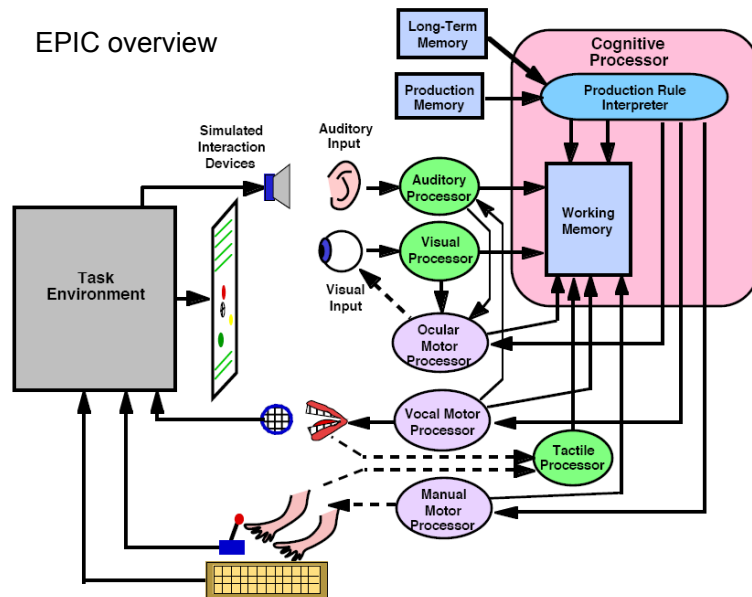
Result: predicted behavior and time course for that scenario and task strategy

Needs comprehensive psychological theory, quite complex; used mostly in a research settings

Example: EPIC Architecture

- Developed to represent executive processes that control other processes during *multiple task* performance
- **Executive-Process Interactive Control** (Kieras & Meyer, mid-1990s)
- Basic assumptions
 - Production-rule cognitive processor
 - Parallel perceptual and motor processors
- Fixed architectural properties
 - Components, pathways, and most time parameters
- Task-dependent properties
 - Cognitive processor production rules (strategy)
 - Perceptual recoding
 - Response requirements and styles

EPIC overview



GOMS (Card, Moran, & Newell, 1983)

- Model-based methodology based on simplified cognitive architectures
- An approach to describing the knowledge of *procedures* that a user must have in order to operate a system
 - **Goals** - what goals can be accomplished with the system
 - **Operators** - what basic actions can be performed
 - **Methods** - what sequences of operators can be used to accomplish each goal
 - **Selection Rules** - which method should be used to accomplish a goal
- Well worked out, practical, but limited due to simplifications
- Often in the "sweet spot" - lots of value for modest modeling effort



Keystroke-level model

1. Choose one or more representative task **scenarios**
2. Have **design** specified to the point that keystroke-level actions can be listed.
3. List the **keystroke-level actions** (operators) involved in doing the task.
4. Insert **mental operators** for when user has to stop and think.
5. Look up the standard **execution time** to each operator.
6. **Add** up the execution times for the operators.
7. The total is the **estimated time** to complete the task (sum of times for tasks t_i multiplied by frequency n_i)

$$T_{execute} = \sum_i t_i * n_i$$

KLM – operators and times

K - Keystroke (0.12 - 1.2 sec; 0.28 for ordinary user)

- Pressing a key or button on the keyboard
- Different experience levels have different times
- Pressing SHIFT or CONTROL key is a separate keystroke
- Use type operator T(n) for series of n Ks done as a unit

P - Point with mouse to a target on the display

- Follows Fitts' law if possible: $0.1 * \log_2 (D/S + 0.5)$
- Typically ranges from .8 to 1.5 sec, average (text editing) is 1.1 sec.

B - Press/release mouse button (.1 sec; click is .2).

- Highly practiced, simple reaction

KLM – operators and times

H - Home hands to keyboard or mouse (.4 sec)

W - Wait for system response

- Only when user is idle because can not continue
- Have to estimate from system behavior
- Often essentially zero in modern systems

M - Mental act of thinking

- Represents pauses for routine activity
- New users often pause to remember or verify each step
- Experienced users pause and think only when logically necessary
- Estimates ranges from .6 to 1.35 sec; 1.2 sec is good single value

Example: file deletion in MacOS, original design, experienced user

General procedure: Find file icon and drag into trash can

Assumptions:

- user thinks of selecting+dragging icon as a single operation
- Finding to-be-deleted icon is still required
- Moving icons to the trash can is highly practiced:
 - The trash can does not have to be located, so finding the trash can is overlapped with pointing to it
 - Verifying that trash can has been hit overlaps with pointing to it
 - Final result not checked since it is redundant with verifying that the can has been hit

Operator sequence:

initiate the deletion **M**, find the file icon **M**, point to file icon **P**, press and hold mouse button **B**, drag file icon to trash can icon **P**, release mouse button **B**, point to original window **P**

□ **Total time = 3P + 2B + 2M = 5.9 sec**

Example: command key file deletion, experienced user

General procedure: select file icon and hit a command key

Assumptions:

- User operates both mouse + key with right hand
- Right hand starts and ends on the mouse

Operator sequence: initiate the deletion **M**, find the icon for the to-be-deleted file **M**, point to file icon **P**, click mouse button **BB**, move hand to keyboard **H**, hit command key **KK**, move hand back to mouse **H**

- Total **time = P + 2B + 2H + 2K + 2M = 5.06 sec**

Only slightly faster, due to the need to move the hand!

Other models in GOMS family

- Critical-Path Method GOMS (CPM-GOMS)
 - Express activities in terms of Model Human Processor → task network → analyze for critical path
- Natural GOMS Language (NGOMSL)/Cognitive Complexity Theory (CCL)
 - basic GOMS concept as simple production system
 - hierarchical actions as sequential/hierarchical rules, eventually keystroke level operators
- Executable GOMS Language (GOMSL)/GLEAN
 - Formalized and executable version of NGOMSL.
 - *GLEAN* - a simplified version of the EPIC simulation system (**GOMS Language Evaluation and Analysis**)

Model-based vs. inspection evaluation

| | <i>Cognitive walkthrough</i> | <i>Heuristic evaluation</i> | <i>Model-based</i> |
|--------------------|------------------------------|-----------------------------|--------------------|
| <i>Stage</i> | Throughout | Throughout | Design |
| <i>Style</i> | Lab | Lab | Lab |
| <i>Objective?</i> | No | No | Somewhat |
| <i>Measure</i> | Qualitative | Qualitative | Qual. & Quan. |
| <i>Information</i> | Low level | High level | Low level |
| <i>Immediacy</i> | N/A | N/A | N/A |
| <i>Intrusive?</i> | No | No | No |
| <i>Time</i> | Medium | Low | Medium |
| <i>Equipment</i> | Low | Low | Low |
| <i>Expertise</i> | High | Medium | High |