Evaluationsmethoden

Ilhan Aslan, Chi Tai Dang, Björn Bittner, Katrin Janowski, Elisabeth André

Human Centered Multimedia
Institute of Computer Science
Augsburg University
Universitätsstr. 6a
86159 Augsburg, Germany

Wiederholung –
Evaluationstechniken

- Zwei Arten von Evaluationstechniken:
  - **Empirische Evaluation** mit Nutzern
  - **Analytische Evaluation** mit (Usability-) Experten
- Sowohl **summativ**, als auch **formativ** durchführbar
- Können aufeinanderfolgend durchgeführt werden
- Beispiel:
  1. Analytische Evaluation: Erkennen und Beheben von offensichtlichen, großen Usability-Problemen
  2. Empirische Evaluation : Erkennen und Beheben weiterer nutzerspezifischer Usability-Probleme

- **Empirische Evaluationen (= Nutzerstudien)**
  - Arten von empirischen Evaluationen
    - **Befragungstechniken:** Befragung der Nutzer nach/vor/während der Benutzung des Systems (z.B. Interviews, Fragebogen)
    - **Beobachtungstechniken**: Aufzeichnung der Benutzung des Systems durch den Nutzer im Feld, Labor oder mittels Simulationen (z.B. Videoaufzeichnung, Protokolle)

- **Empirische Evaluationen**
  - **Befragungstechniken:**
    - Meinung der Nutzer wird abgeklärt
      - ➢ Subjektive Ergebnisse
      - ➢ in erster Linie qualitative Messwerte
    - Quantitative Messwerte durch:
      - geschlossene Fragen
      - Annotation anhand eines Analyseschemas für qualitative Messwerte (z.B. offene Fragen)

Empirische Evaluation: Beobachtungstechniken

Stephan Hammer, Ilhan Aslan,
Andreas Seiderer, Simon Flutura
Elisabeth André

Human Centered Multimedia
Institute of Computer Science
Augsburg University
Universitätsstr. 6a
86159 Augsburg, Germany

Beobachtungstechniken

- Liefern objektive Daten über den Nutzer
  - Verhalten
  - Vorlieben
  - in unterschiedlichen Situationen
- Unterscheidungsmerkmale
  - Was wird beobachtet?
  - Wo wird beobachtet?
  - Wie wird beobachtet?
  - Wie wird ausgewertet?

Was wird beobachtet?

- Nutzer **ohne** das System (Anforderungsanalyse)
  - Wer ist der Nutzer (Nutzerspezifikation)?
  - Welche Aufgaben hat der Nutzer zu erledigen (Taskspezifikation)?
  - In welchem Umfeld agiert der Nutzer (Kontextspezifikation)?
- Nutzer **mit** dem System (Nutzerzentriertes Prototyping)
  - Wie kommt der Nutzer mit dem System klar?
  - Welche Probleme existieren und warum?
  - Wo gibt es Potential für Verbesserungen?
  - Welche Aufgaben fehlen / sind unnötig?
  - Sind alle Aktionen sichtbar / verfügbar?
  - Ist das Feedback angemessen? …

Was wird beobachtet? – Nutzer mit System

- Freie Nutzung
  - Nutzer kann System und dessen Funktionen frei nutzen.
  - Oft am Ende des Entwicklungsstadiums
    - Hohe Funktionalität der Prototypen
    - High-Fidelity Prototyp
    - Finales Produkt
  - Fragen:
    - Wie arbeitet der Nutzer mit dem System?
    - Welche Funktionen bevorzugt der Nutzer?
    - Wo gibt es Probleme?...

Was wird beobachtet? – Nutzer mit System

- Abarbeitung von vorgegebenen Aufgaben (Tasks)
  - Nutzern werden verschiedene Tasks vorgegeben und durchlaufen diese
  - Oft am Anfang und der Mitte des Entwicklungsstadiums
    - Low und High-Fidelity Prototyp
    - Horizontale und in erster Linie vertikale Prototypen
  - Fragen:
    - Ist eine Optimierung der untersuchten Tasks möglich?
    - Sind alle Aktionen vorhanden / sichtbar?
    - Passt das Feedback?

Methode des Lauten Denkens

- Sehr oft bei der Abarbeitung von vorgegebenen Aufgaben
- Manchmal bei der freien Nutzung
- Nutzer verbalisiert seine Gedanken während der Nutzung
  - Was denkt er über das UI?
  - Welche Erwartungen hat er und werden diese erfüllt?
  - Ist er zufrieden oder verwirrt?
- **Vorteile:**
  - Sehr gute und wichtige Informationen
- **Nachteile:**
  - Nicht natürlich
  - Kann ungewohnt sein
  - Nutzer vergisst seine Gedanken zu verbalisieren

Wo wird beobachtet?

Wo wird der Nutzer aufgezeichnet?
(Analysephase und nutzerzentriertes Prototyping)

- Direkte Beobachtung:
  - im natürlichen Umfeld des Nutzers (Feldstudie oder In-situ)
    - Ethnographie
  - in einer kontrollierten Umgebung (Laborstudie)
    - Reale Simulationen
    - Virtuelle Simulationen
    - Hybride Simulationen
- Indirekte Beobachtung (Tracking von Nutzeraktivitäten)

Wo wird beobachtet? - Feldstudie

- Nutzer interagiert in seiner natürlichen Umgebung während er beobachtet wird (z.B. im Büro, zu Hause oder im Kaufhaus) (Big Brother-Prinzip)
- Gerade in der Anforderungsanalyse sehr wichtige Methodik zur Nutzer-, Task- und Kontextanalyse

Wo wird beobachtet? - Feldstudie

Universität Augsburg University

Vorteile:

- Evaluation findet unter realistischen Bedingungen statt
  - Unbewusstere Beobachtung
  - Nutzer passt Verhalten oft nicht an die Situation an
  - Sehr natürliche Aktivitäten des Nutzers
  - Sehr hochwertige Daten
- Kontext vorhanden, dadurch können situationsspezifische Usability-Probleme eher erkannt werden.
- Evaluation kann kontinuierlich über längeren Nutzungszeitraum durchgeführt werden

Wo wird beobachtet? - Feldstudie

- Probleme:
  - zeitaufwändig und kostspielig
  - Ergebnisse nicht ohne weiteres reproduzierbar wegen nicht kontrollierbarer Kontextbedingungen (Geräusche, Störungen...)
  - Studie **mit System**
    - Oft schwer durchzuführen (z.B. bei mobilen Applikationen)
    - Meist nur spät im Entwicklungsprozess mit robusten Prototypen möglich
- Mit System sinnvoll, wenn:
  1. Kontext eine wichtige Rolle spielt
  2. robuster Prototyp vorhanden und/oder
  3. Langzeitstudie erforderlich ist.

Wo wird beobachtet? - Feldstudie

Universität Augsburg University

Ethnographie:

- Technik, die Beobachtungen im Feld aufzeichnet
- Untersuchung des Nutzers in seiner Umgebung, um seine Aufgaben und seine Vorlieben herauszufinden
  - Analyse der Nutzerbedürfnisse
  - Bezug zur Taskanalyse
    (Wie werden welche Aufgaben bisher ausgeführt?)
  - Kontextinformationen
    (Welcher Kontext fließt in die Nutzung des Systems ein?)

Wo wird beobachtet? - Feldstudie

Universität Augsburg University

Ethnographie – Vorgehen:

- Kombination von Beobachtung und Interview!
- Vorbereitung eines Interviews
  - Was will man vom Nutzer wissen?
- Nutzer während seiner Arbeit aufzeichnen und später analysieren
- Beim Interview Fragen zu Beobachtungen um Missverständnisse zu vermeiden

- **Achtung:** Analyse von Videomaterial ist sehr aufwendig und manchmal nicht das beste Mittel
- Methode des lauten Denkens evtl. effizienter

Wo wird beobachtet? - Laborstudie

- Funktion
  - Forschung im Bereich der angewandten Kognitionswissenschaft und der HCI
  - Untersuchung und Entwicklung von Gestaltungskonzepten für interaktive Software und andere Formen der HCI
  - Im Labor werden Nutzungsbedingungen simuliert, um mit Endnutzern Tests durchzuführen
  - Nutzer erhalten oft konkrete Aufgaben und werden dann bei der Bearbeitung beobachtet und „gemessen“ (Zeit, Fehler usw.)
  - Durchführung von Benutzertests, Evaluationsuntersuchungen, Anforderungsanalysen etc.

Wo wird beobachtet? - Laborstudie

- Arten von Laborstudien (Simulationsart):
  - Reale Simulationen
  - Virtuelle Simulationen
  - Hybride Simulationen

Wo wird beobachtet? - Laborstudie

Universität Augsburg University

Reale Simulation:

- Im Labor wird die reale Umgebung des Nutzers nachgebildet und mit Aufzeichnungstechniken instrumentiert (z.B. mit Kameras)
- Nutzer interagiert mit der realen bzw. eigentlichen Umgebung
- z.B. iDorm2 (intelligente Wohnung)

http://iieg.essex.ac.uk/idorm2/index.htm

Wo wird beobachtet? - Laborstudie

Universität Augsburg University

Reale Simulation – Beispiel:

- Philips HomeLab
  - ganzes Haus (Wohn-, Kinder-, und Schlafzimmer, Küche, Bad und Diele) zum Usability Lab umgebaut. (ähnlich zum „Big Brother“-Container)
  - erlaubt die Beobachtung der Nutzung neuer Technologien unter realistischen Bedingungen

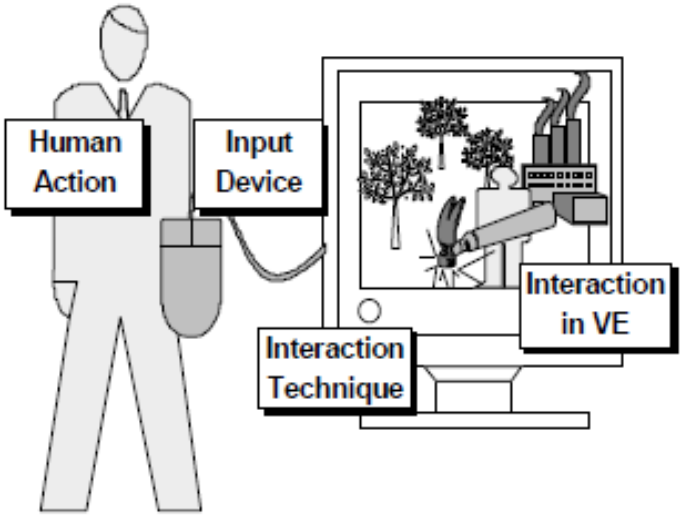
http://www.noldus.com/default/philips-homelab

Wo wird beobachtet? - Laborstudie

Universität Augsburg University

Virtuelle Simulation:

- Reale Simulation sehr teuer und zeitaufwendig
- Virtuelle Simulation simuliert die reale Umgebung in einer virtuellen Welt (z.B. virtuelles Augsburg).
- Nutzer interagiert mit den Objekten der virtuellen Simulation
- Beispiel: **Ubiwise Simulator for Ubiquitous Computing**



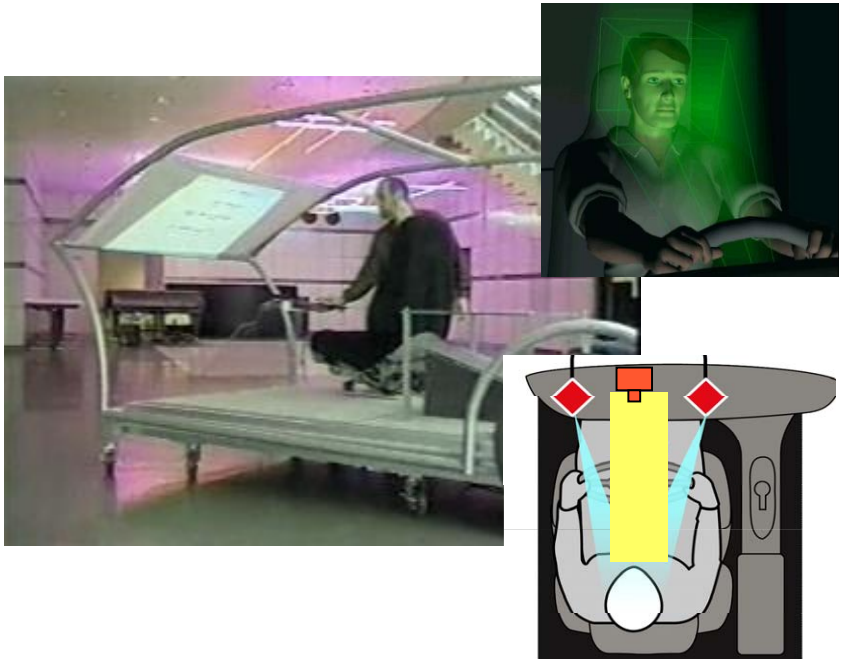
http://home.comcast.net/~johnjbarton/ubicomp/ur/ubiwise/index.htm

Wo wird beobachtet? - Laborstudie

Universität Augsburg University

Hybride Simulation:

- Kombination der realen mit der virtuellen Simulation
  - Reale Interaktionsgeräte (z.B. Mobiltelefone oder Cockpit)
  - Virtuelle Simulation der Umgebung
- Einsparung von Kosten und Zeit
- Realistischere Studien als reine virtuelle Simulationen



Wo wird beobachtet? - Laborstudie

- Vorteile:
  - Durchführung von Experimenten unter kontrollierbaren Bedingungen
    - ermöglicht Reproduzierbarkeit und Vergleichbarkeit von Ergebnissen
  - Leichter, schneller und billiger durchführbar als Feldtests

- Probleme:
  - kein realer Kontext
  - Usability Lab mit Spezialausstattung und Experten notwendig um realtische Studien und wertvolle Daten zu erhalten

Wo wird beobachtet? - Durchführung von Feld- und Laborstudie

- Lokale Studie
  - Keine räumliche und zeitliche Trennung des Evaluators und der Versuchsperson (=> Sitzen in einem Raum)
  - Keine Nutzung von Technologie zur Übertragung für einen räumlich getrennten Evaluator

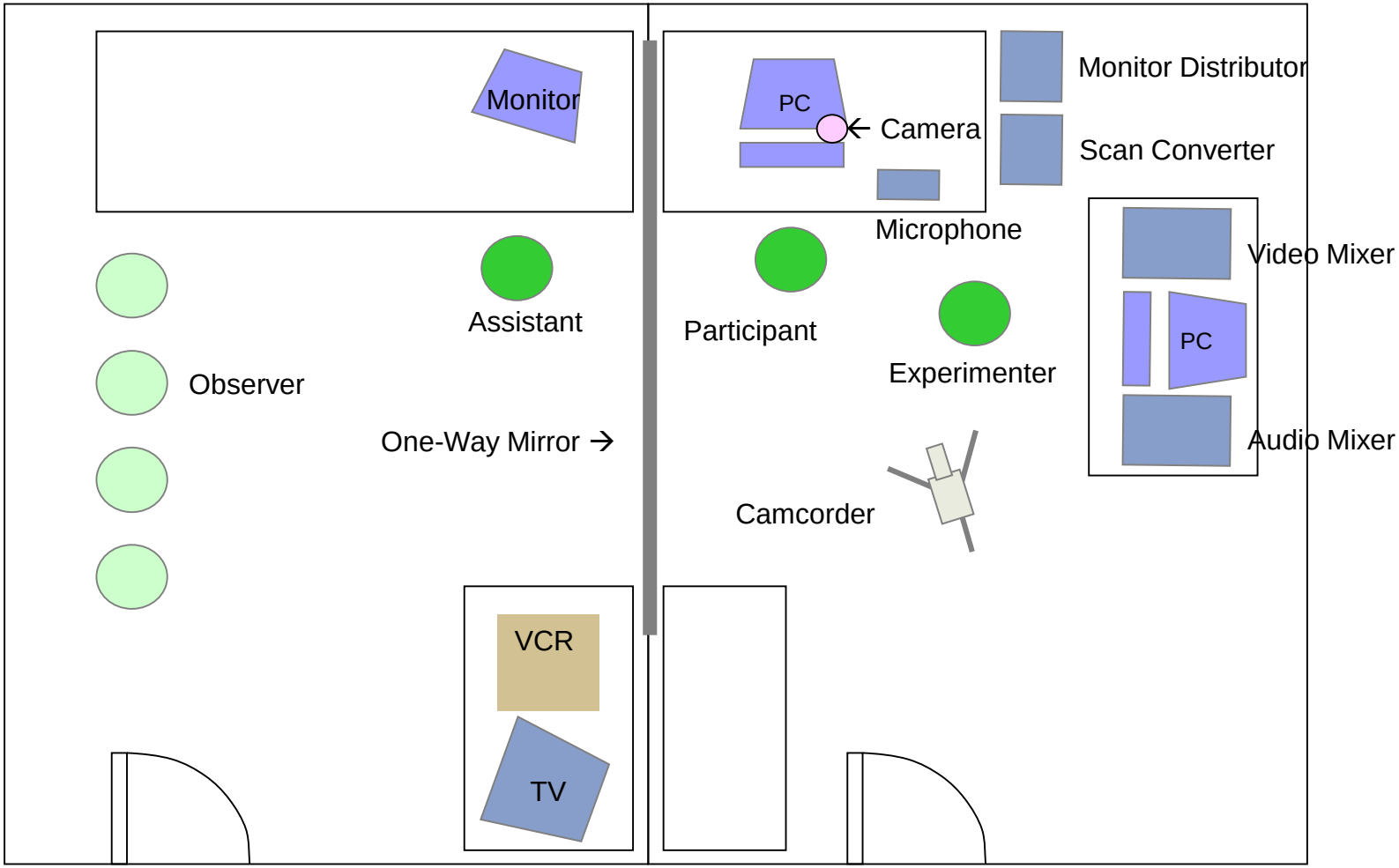
- Entfernte Studie
  - Räumliche und/oder zeitliche Trennung des Evaluators und der Versuchsperson
  - Nutzung von Technologie zur Übertragung für einen räumlich getrennten Evaluator
    - Visuelle Technologien (Kamera)
    - Logging der Nutzeraktivitäten

Wo wird beobachtet? - Durchführung von Feld- und Laborstudie

Universität Augsburg University

Räumlich entfernte Laborstudien

Usability-Labor (Schema)



Wo wird beobachtet? - Durchführung von Feld- und Laborstudie

Universität Augsburg University

Räumlich entfernte Laborstudien






Wo wird beobachtet? - Durchführung von Feld- und Laborstudie

Universität Augsburg University

Räumlich entfernte Laborstudien

Trackingstudie:

- Einsatz von Aufzeichnungstechnik zum automatischen Loggen der Nutzeraktivitäten (z.B. Mausklicks, Verweildauer)
- Nutzer interagieren in einem komplett natürlichen Umfeld (Webseite von zu Hause)
- Viele Nutzer wissen nichts von der Aufzeichnung

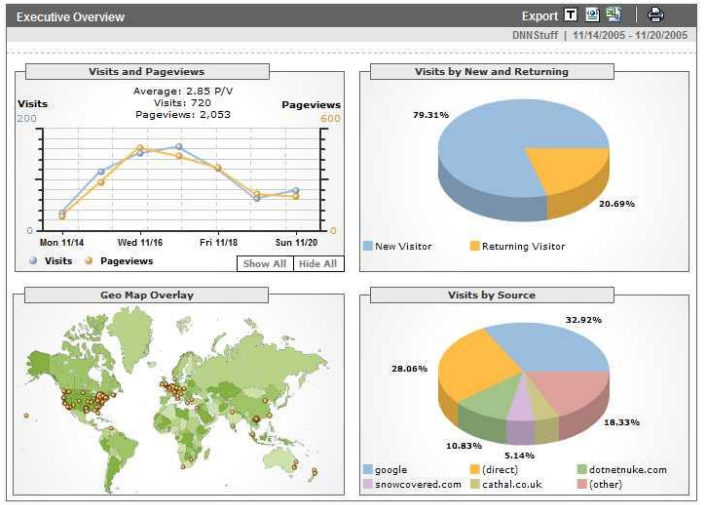
➢ Sehr natürliches Verhalten der Nutzer

➢ Primär quantitative Daten

  – Statistische Analyse

  – Beispiel: Google Analytics

  http://www.google.com/analytics/



Wo wird beobachtet? - Durchführung von Feld- und Laborstudie

Universität Augsburg University

Räumlich entfernte Laborstudien

Trackingstudie:

- Vorteile:
  - Sehr schnell und sehr einfach
  - viele Daten von vielen Nutzern
  - sehr realistische Daten
- Nachteile:
  - Schwer zu interpretierende Daten
    - Warum klicken die Nutzer häufig einen Link auf einer Webseite?
    - Warum haben die Nutzer einen Vorgang abgebrochen?
  - Was sind mögliche Lösungen / Änderungswünsche?

Wo wird beobachtet? – Fazit

- **Feldstudien am Besten um sehr realistische Daten zu erhalten. Diese sind aber sehr aufwendig und teuer durchzuführen**.
  - Oft in der Anforderungsanalyse und am Ende des nutzerzentrierten Prototypings
- **Laborstudien sind sehr kontrolliert, weniger teuer und leichter durchzuführen als Feldstudien, aber auch weniger realistisch.**
  - Während des kompletten nutzerzentrierten Prototypings
- **Trackingstudien bieten am schnellsten und leichtesten eine Vielzahl an Log-Daten. Diese sind aber am schwierigsten interpretierbar.**
  - In erster Linie in der Anforderungsanalyse von alten Systemen
  - Zum Test von neuen Webinterfaces

➢ Jede Studienart ist sinnvoll in unterschiedlichen Phasen mit unterschiedlichen Zielen der Evaluatoren.

Wie wird beobachtet?

- Verschiedene Aufzeichnungstechniken:
  - **Text / Notizen**
    - einfach, billig, spontan praktikabel
    - Aber: Details können verloren gehen (Kontexte)
  - **Bilder**
    - Schlüsselbilder von interessanten Momenten
    - Aber: Details können verloren gehen (Kontexte)
  - **Audio**
    - Gut für Methode des lauten Denkens
    - Aber: Details können verloren gehen (Kontexte)

Wie wird beobachtet?

- Verschiedene Aufzeichnungstechniken:
  - **Video**
    - Am besten mit mehreren Kameras
      1. Sicht auf den Probanden
      2. Sicht auf die Umgebung (Kontext)
      3. Sicht des Probanden
  - **Physiologische Datenaufzeichnung**
    - Nutzeraktivität, -emotionen und -zustand in Form von Puls, Hautleitwert, Atmung etc.
  - **Eyetracker**
    - Erfasst die Blickbewegung des Nutzers
  - **(Software-)Logging**
    - Details der Nutzerinteraktion mit Software (z.B. Mausklicks)

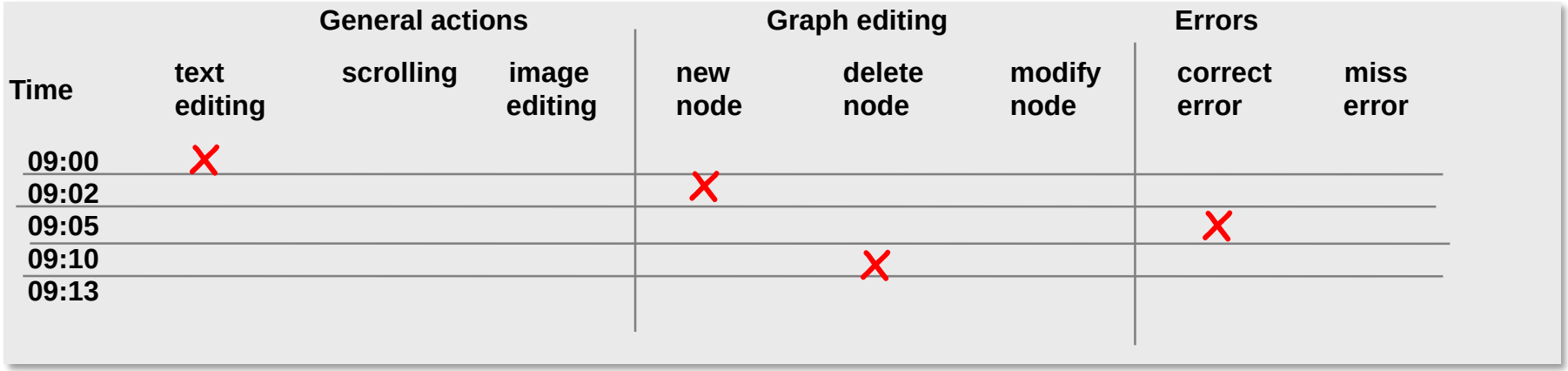
Wie wird beobachtet?

Wer zeichnet auf?

- Evaluator
  - Klassische Beobachtungsstudie
  - Typischerweise lokale Studien im Feld oder **Labor (primär)**
  - ➢ Evaluator und Versuchsperson sind zeitlich nicht getrennt.
  - ➢ Es findet eine räumliche Trennung statt, falls der Versuchsleiter im Nebenzimmer sitzt!
- Probanden
  - Diary Studies (z.B. Cultural Probes und ESM)
  - Typischerweise entfernte Studien im **Feld (primär)** oder Labor
  - ➢ Evaluator und Versuchsperson sind räumlich getrennt.
  - ➢ Synchron, wenn sie zeitlich nicht getrennt sind und asynchron, wenn sie auch zeitlich getrennt sind.

Wie wird beobachtet? – Aufzeichnung durch den Evaluator

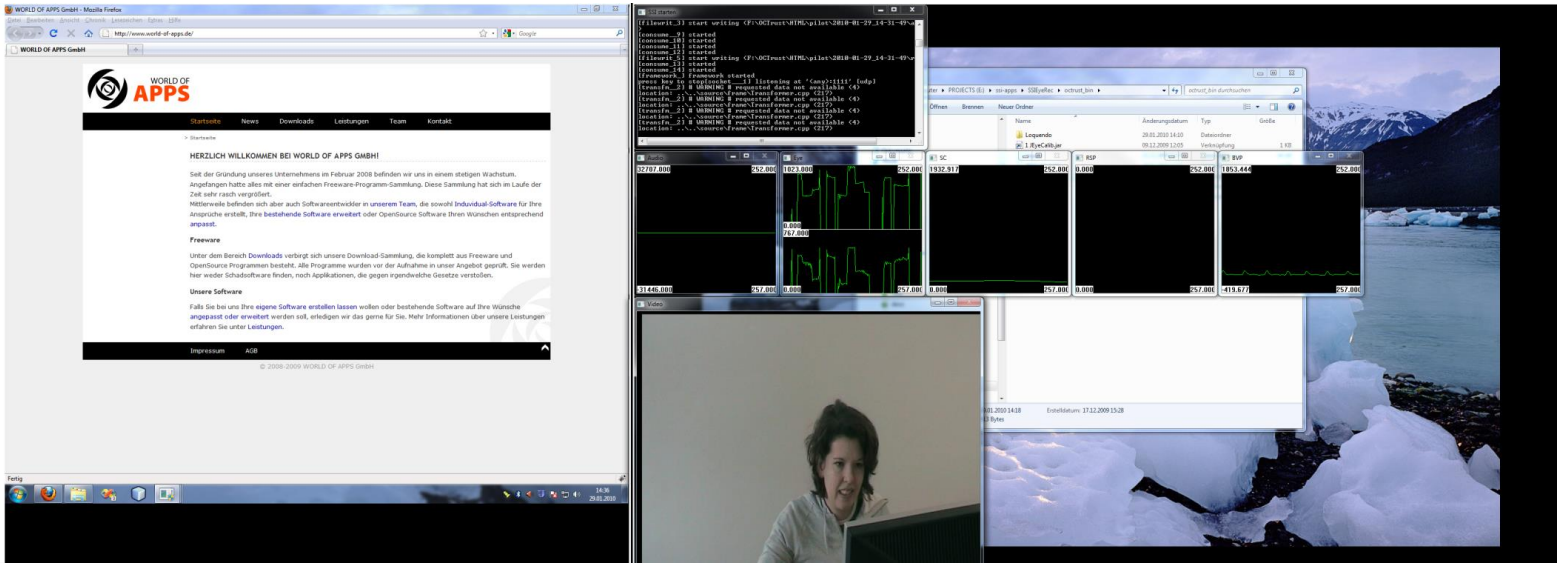
- Handschriftliche Textaufzeichnung:
- Problem:
  - unterschiedliche Daten (Kommentare, Systemeingaben, Gesten, Mimik etc.) müssen zeitlich korrekt angeordnet werden.
- Ansatz:
  - „Musical Scores“ (vgl. Notation der zeitlichen Notenreihenfolge eines mehrstimmigen Musikstücks)



Wie wird beobachtet? – Aufzeichnung durch den Evaluator

Universität Augsburg University

- Video und Audioaufzeichnung:
  - Aufzeichnung vieler verschiedener Sichten – Beispiele:
    - Aufzeichnung des Bildschirms des Nutzers
    - Zwei Kameras (Blickrichtung des Nutzers + komplette Szene)
  - Anschließende Diskussion mit der Testperson
    - Warum hast du dies und jenes gemacht?
    - Was hast du da versucht?

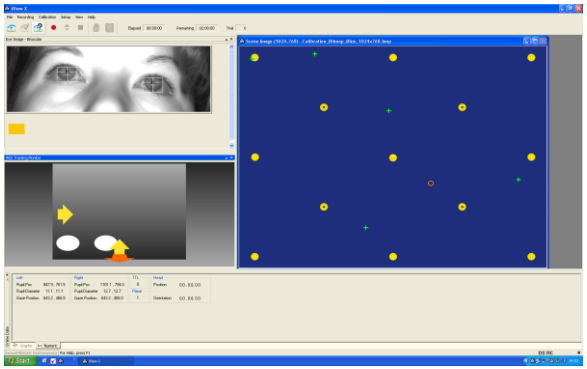


Wie wird beobachtet? – Aufzeichnung durch den Evaluator

- (1) Testperson, (2) Versuchsleiter (Evaluator), (3) Sensoren (Puls, Hautleitwert, Atmung), (4) Eyetracker, (5) Kamera
- Verwendung von SSI zum synchronen Aufzeichnen von Nutzerstudien (http://openssi.net/)

Wie wird beobachtet? – Aufzeichnung durch den Evaluator

- Eye Tracking



SMI

Pupil (Pupil Labs)

The Eye Tribe (Oculus)

Wie wird beobachtet? – Aufzeichnung durch den Probanden

- Experience Sampling Method (ESM)
  - Beantwortung von kurzen Fragebögen:
    - auf Anfrage (Alarm, SMS, Zeitplan)
    - immer sofort in der entsprechenden Situation
  - Dokumentation von Verhalten, Gefühlen und/oder Gedanken
  - Aufzeichnung und Beantwortung mit z.B. Stift + Papier und/oder Kamera
  - **Vorteil:**
    - Einblick in Situationen, die von Designern schlecht beobachtbar sind (z.B. Essverhalten)

Wie wird beobachtet? – Aufzeichnung durch den Probanden

- Experience Sampling Method (ESM)

- Klassischerweise **entfernte Studien:**
  - Ferngesteuertes **synchrones** Testen:
    - Durchführung in Echtzeit, aber Testmonitor räumlich getrennt von den Versuchsteilnehmern
  - Ferngesteuertes **asynchrones** Testen:
    - Testmonitor räumlich und zeitlich getrennt
- Typischerweise 30 bis 80 Teilnehmer, die über einen Zeitraum von ein bis drei Wochen ca. 10 Erinnerungen pro Tag erhalten
- Teilnahme der Forscher beschränkt sich auf Interviews am Anfang und am Ende einer Studie.

Wie wird beobachtet? – Aufzeichnung durch den Probanden

Universität Augsburg University

Experience Sampling Method (ESM) - Alarmmechanismen

Kategorien	Beschreibung	Nachteile
Arten des Alarms	Zufällig	Flexibles Werkzeug benötigt
	Nach Zeitplan	Könnte zu höherer kognitiver Belastung führen
	Ereignisbasiert	Könnte zu höherer kognitiver Belastung führen, wenn der Alarm vom Teilnehmer selbst ausgelöst wird
Art der zeitlichen Planung	Alarmsignale nur zu bestimmten Tageszeiten	Flexibles Werkzeug benötigt, interessante Situationen könnten verpasst werden
	Anzahl der Alarmsignale pro Tag	Flexibles Werkzeug benötigt, Ergebnisse könnten verfälscht werden, wenn Teilnehmer weiß, wie viele Signale er pro Tag erhält
	Anzahl der Alarmsignale insgesamt	Flexibles Werkzeug benötigt
Mitteilungs-mechanismus	Auditiv	Könnte in bestimmten Situationen inadäquat sein (Theater, Kino)
	Taktil	Wird u.U. nicht wahrgenommen

Wie wird beobachtet? – Aufzeichnung durch den Probanden

- Cultural Probes:
  - Nutzer erhalten ein Paket, mit dessen Inhalt (z.B. Kamera) sie unterschiedliche Aufgaben erfüllen sollen
  - Beispiele:
    - Mache Fotos von Deiner Umgebung
    - Markiere Fotos mit Stickern, um die Beziehungen zu Arbeitskollegen zu veranschaulichen
  - Probleme:
    - Ergebnisse schwer zu interpretieren, daher eher zur Inspiration geeignet.
    - Recht hoher Aufwand für Nutzer

Wie wird beobachtet? – Aufzeichnung durch den Probanden

- Framework „EDDY“:
  - Sammeln von Daten mittels einer einzigen Smartphoneanwendung
  - **Aktive „Sensoren“:** Texteingabe, Foto, Video-, Audioaufnahme, RFID
  - **Passive Sensoren:** GPS, EKG, 3D-Beschleunigung
  - Zusammengehörende Daten werden in Paketen gespeichert
  - ➢ Zusammenhang zwischen einzelnen Daten kann später wieder erkannt werden
  - Beispiel: Datenpaket

Titel:
„Mittwoch – Mittagessen“

Text:
„Schnitzel mit Kartoffel-salat und Beilagensalat“

Wie wird beobachtet? – Aufzeichnung durch den Probanden

- Framework „EDDY“:
  - Bei digitaler ESM und Cultural Probes einsetzbar
  - **Vorteile:**
    - Smartphone ersetzt komplettes Paket mit Kamera, Stift + Papier usw.
      - ➢ weniger Aufwand für Probanden
    - Zusammengehörende Daten werden zusammen gespeichert
      - ➢ Erleichterung der Reflexion und Auswertung
    - ➢ Steigerung der Datenvielfalt
    - ➢ Motivation der Probanden

Wie wertet man aus?

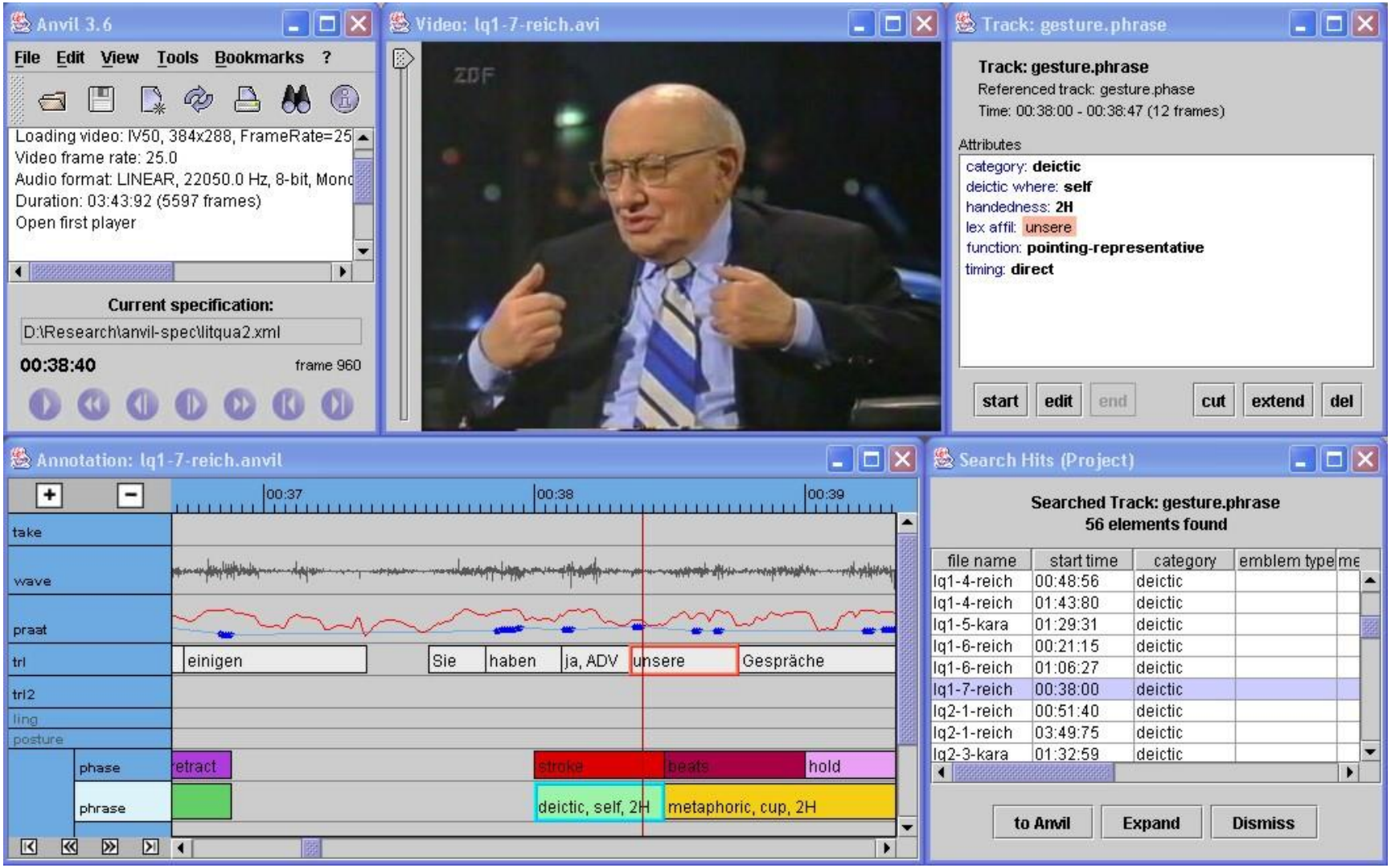
- Quantitative Daten kann man direkt analysieren
  (siehe Foliensätze zur Datenanalyse)
- Wie analysiert man qualitative Daten?
  - Befragungstechniken: offene Fragen
  - Beobachtungstechniken:
    Text, Audio-, Videoaufnahmen, Physiologische Daten, Blickdaten

➢ Annotation

  - Auswertung der objektiven Daten und Zuweisung einer messbaren Bedeutung
  - Nutzung von sogenannten Analyseschemas
    (z.B. Zeitpunkte an denen der Nutzer den rechten Arm hebt)
  - Nach der Annotation messen / zählen der Annotationen

Universität Augsburg University

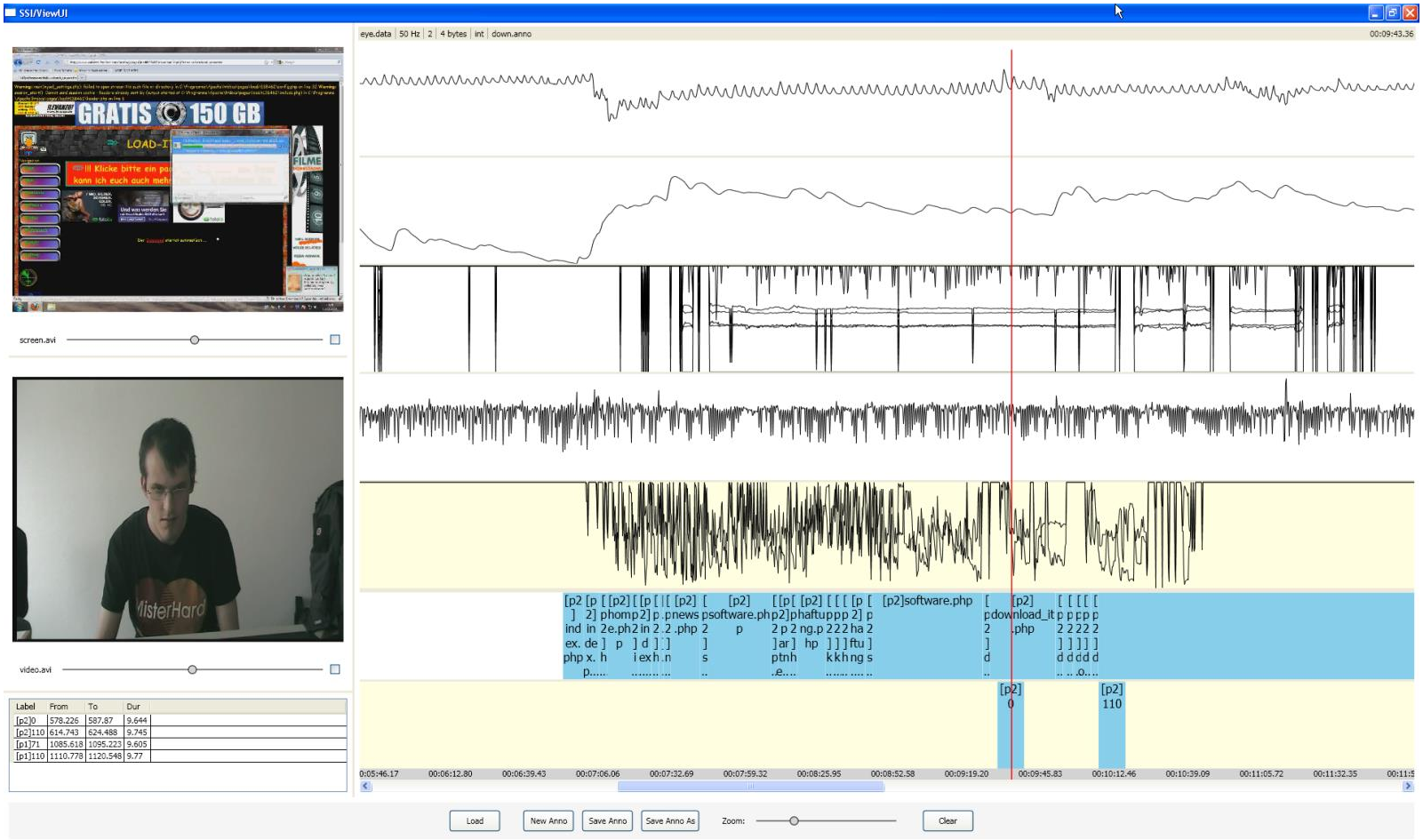
Verwendung von Annotationswerkzeugen (Anvil)



Wie wertet man aus?

Verwendung von Annotationswerkzeugen (ModelUI von SSI)

– Analyse physiologischer Daten, Video / Audiodaten und Blickdaten



Wie wertet man aus?

- Probleme der Annotation
  - Ist das Analyseschema angemessen?
  - Sind alle interessanten Aspekte enthalten?
  - Mögliche Fehler durch falsche Interpretation der Aktivitäten
    - Mehrere Personen für Annotationen nötig
    - Überprüfung der Ergebnisse mittels Κ (Kappa) – Berechnung (Inter-rater reliability, Concordance)
  - Sehr zeitaufwendig und teuer

Auswertung durch deskriptive Statistik

- Quantitative Analyse
  - Numerische Methoden zur Ermittlung von Summen, Größen etc.
- Beispiele:

37	3	2	1	1	4

  - **Mode:** Zahl, die am häufigsten auftritt: **1**
  - **Mean / Average:** Durchschnitt bzw. arithmetisches Mittel: 48 / 6 = **8**
  - **Median:** Wert an mittlerer Stelle aus sortierter Liste
    - Robuster gegen Ausreißer als der Durchschnitt
    - Ungerade Anzahl:

1	2	**4**	5	18

→ **4** (Mean:6)

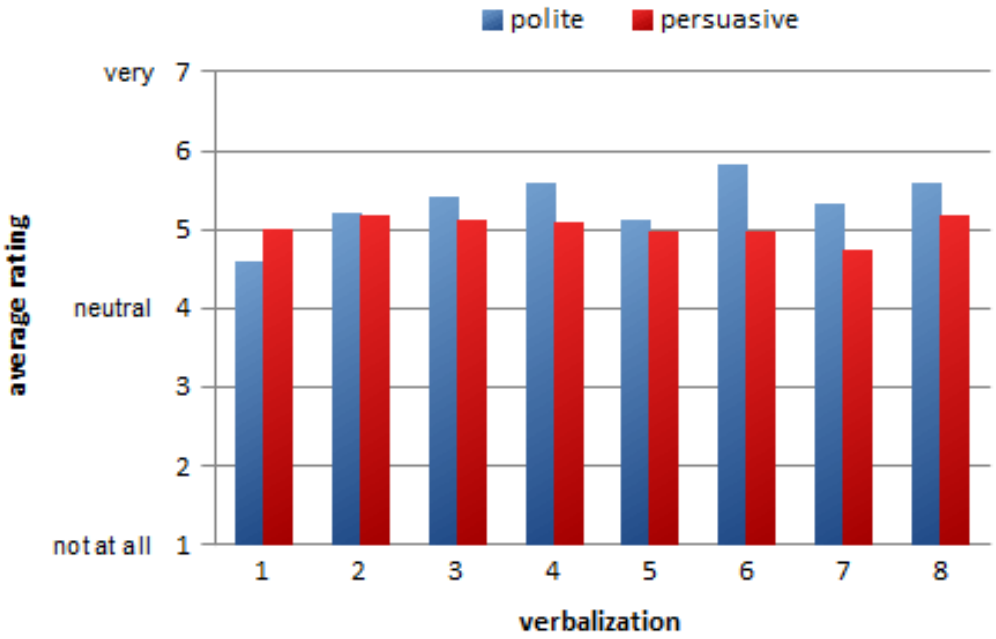
    - Gerade Anzahl:

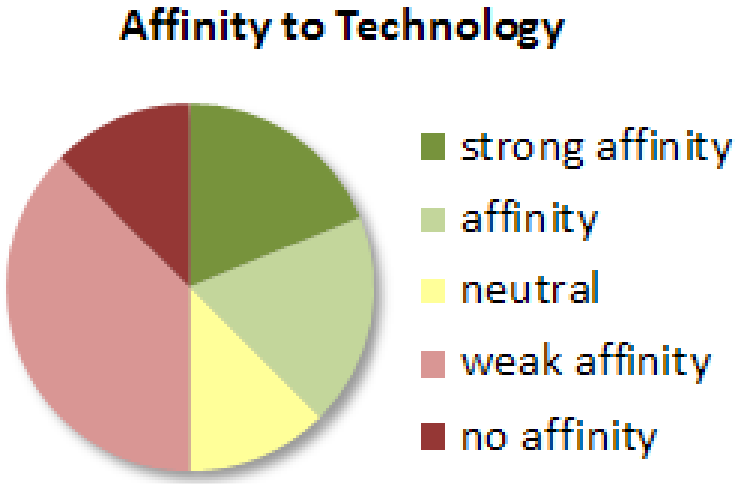
1	1	**2**	**3**	4	37

→ (2+3)/2 = **2,5** (Mean:8)
  - **Prozentsätze**
  - **Signifikanzen** (siehe Foliensätze zur Datenanalyse)

Graphische Repräsentation:

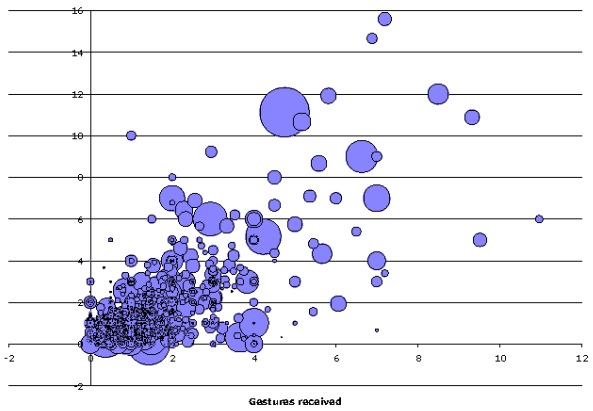
- Gibt einen schnellen Überblick über die Daten
- Je nach Daten passende Darstellung / Graphen verwenden

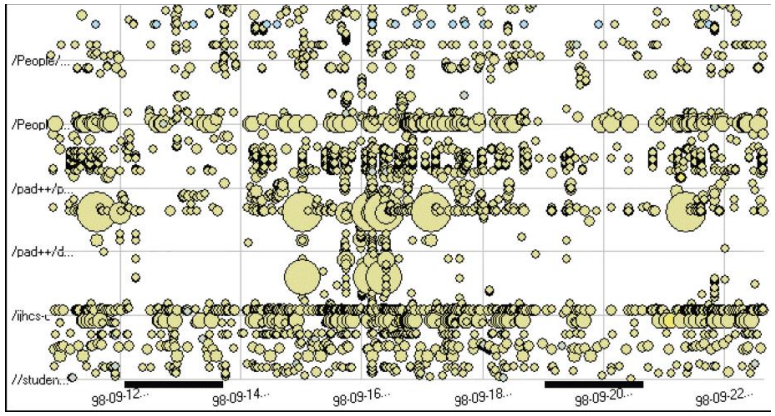




Graphische Repräsentation:

Visualisierung von Logfiles





Heatmaps für mit einem Eyetracker gesammelte Blickdaten