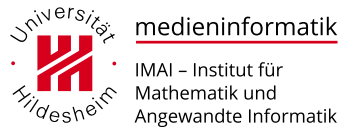


Modalität, Codalität, Realität

Jörg Cassens

Medieninformatik

WS 2019/2020



Einleitung

- Medieninformatik ist eine junge Disziplin
- Viele der digitalen Medien, die wir heute kennen, gibt es erst seit wenigen Jahren oder Jahrzehnten
- Neue Möglichkeiten, aber auch neue Herausforderungen
- Hier betrachten wir vor allem Systeme, die die Möglichkeiten der Interaktion durch koordinierte und gleichzeitige Nutzung mehrerer Kanäle erweitern
 - Kombination mehrerer Modalitäten/Codalitäten
 - Verarbeitung natürlicher Sprache
- Immer kleinere und vernetzte Computer ermöglichen die Nutzung von Informationen und Medien immer und überall
- Ambient Computing: Durchdringung des Alltagslebens
 - Kontextualisierte Anwendungen
- Virtual, Augmented und Mixed Reality als aktuelle immersive Technologien

Lernziele

- Entwicklungen im Bereich Digitaler Medien, die über klassische Medientypen hinausgehen, veranschaulichen
- Techniken und Ansätze skizzieren und an Beispielen erläutern
- In die Lage versetzen, neue Trends einzuschätzen
- Vermittlung, daß in der Zukunft neue digitale Medien realisierbar sein werden, die deutlich komplexer, vielseitiger und leistungsfähiger sind als heutige Systeme

1 Medialität, Codalität und Modalität

Ebenen

- **Präsentation & Aufnahme**
 - Die “technische Seite”
 - Hilfsmittel zur Ein- und Ausgabe von Informationen
 - Mögliche Fragen:
 - * Welche Aufnahmegeräte werden gebraucht (Mikrofon)?

* Welche Wiedergabegeräte werden benutzt (Monitor)?

- **Codierung**

- Die Repräsentation
- Form, in der die Information beschrieben wird
- Mögliche Frage:
 - * Werden Texte oder Graphiken benutzt?

- **Wahrnehmung & Produktion**

- Die “menschliche Seite”
- Der genutzte/angesprochene Sinneskanal
- Mögliche Fragen:
 - * Wie wird die Information wahrgenommen (Augen, Ohren)?
 - * Wie wird die Information ausgedrückt (Sprache, Gestik)?

Rekapitulation: Kanäle, Codecs und Medien

Ebenen: Beispiele

- **Präsentation & Aufnahme – Medial**

- Radio: monomedial
- Videokamera: multimedial

- **Codierung – Codal**

- Nur Text, nur Graphik: monocodal
- Gemischt: multicodal

- **Wahrnehmung & Produktion – Modal**

- Nur die Augen ansprechend: monomodal
- Augen und Ohren ansprechend: multimodal

Problem: Unterschiedliche Belegung in unterschiedlichen Kontexten und Fachgebieten

Rekapitulation: Kanäle, Codecs und Medien

Medialität, Codalität und Modalität

Multimedialität: Systeme, in denen mehrere Medientypen (wie Text, Bilder, Videos) benutzt werden, bezeichnet man als *multimediale* Systeme

Multicodalität: Systeme, in denen die gleiche Informationen unterschiedlich codiert vorliegen, bezeichnet man als *multicodale* Systeme

Multimodalität: Systeme, bei denen zur Eingabe oder Ausgabe gleichzeitig und koordiniert mehrere Sinneskanäle zur Kommunikation benutzt werden, bezeichnet man als *multimodale* Systeme

- Medialität legt Fokus auf die technische Repräsentation
- Codalität betrachtet die semantische Repräsentation
- Modalität bezieht sich auf die menschlichen Sinne

Trennung

- Begriffe oft nicht eindeutig benutzt
- Teilweise werden z.B. verschiedene Formen der Repräsentation von Informationen als Modalitäten betrachtet
 - Tabellen
 - Graphiken
 - Gesprochener Text
- Vor allem: sensorische und Repräsentationsebenen häufig nicht klar getrennt
- Hier: Grundsätzliche Betrachtung von
 - Sprache
 - Gesten
 - Mimik
 - bildlicher Darstellungals Codalitäten, die ggf. über verschiedene sensorische Kanäle transportiert werden

Repräsentation und Sensorik

- Keine eindeutige Festlegung durch das technische Medium
 - Bildliche Darstellung als Vektor- oder Pixelgraphik
 - Geste als Video, Animation oder durch Roboter
- Ein Medium kann unterschiedliche Codalitäten transportieren
 - Pixelbild für Gesicht mit Mimik
 - Pixelbild als Schrift
 - Pixelbild als Geste
- Gerade Modalität und Codalität häufig verwirrend verwendet, multimodale Systeme befassen sich häufig mit Multicodalität

Bedeutung

- Multimodalität/-codalität weniger bekannt als Multimedia, spielt aber große Rolle
 - Menschliche Kommunikation intrinsisch multimodal/multicodal (Halliday)
 - Gesprächssituationen umfassen fast immer Gestik und Mimik neben der gesprochenen Sprache
- Gründe, Multimodalität/-codalität bei digitalen Medien zu nutzen:
 - Unterstützung eines natürlichen Interaktionsstils
 - größerer Kommunikationserfolg durch Redundanz
 - bessere Eignung für beeinträchtigte Nutzergruppen (z.B. Seh- oder Höreinschränkung)
 - robustere Erkennung von Nutzereingaben
- Auch bei der Interaktion mit Computern natürlichen Interaktionsstile nachahmen
 - Mehr Spaß an der Nutzung
 - Schnellere Lernerfolge
 - Höhere Zufriedenheit

Multimodale/-codale Nutzereingaben

- Nutzereingaben können vermittelt werden z.B. über
 - Geschriebener Text, Tastaturkommandos
 - Zeigegesten mit der Maus
 - Spracheingabe durch gesprochene Sprache
 - visuelle Eingaben, die Gesten, Mimik und Körperhaltung erfassen
 - Bio-Sensoren für implizite Eingaben (Pulsfrequenz, Brain-Computer-Interfaces)
 - akustische Analyse (Intonation der Sprache, Umgebungsgeräusche)
 - haptische Eingaben über Knöpfe, Tasten, andere Geräte
 - ...
- Kombination von Sprache und anderen Modalitäten häufig anzutreffen

Video 14.1: Put That There (1980)



Put That There (5:58)

Herausforderungen

- Sprache kombiniert mit Gesten ermöglicht es Benutzern, auf Objekte zu zeigen und zu sprechen
- Idee bei "Put That There": nicht alles mit Sprache ausdrücken müssen
- Einerseits für den Nutzer leicht zu erlernen und bequem
- Andererseits für das System einfacher, Äußerungen wie "dies" und "dahin" zu verstehen als "das rote Schiff" und "nach unten rechts"
- Aber: Herausforderung liegt in der Zuordnung und Erkennung multimodaler Zusammenhänge
- "dies" und "dort" sind Anaphern
 - Bezug auf nicht explizit genannten Objekte
- Herausforderung also, die Gesten und Sprechakte den richtigen Objekten zuzuordnen, da beide nicht immer gleichzeitig auftreten

Spracherkennung

- Spracherkennung wird benötigt, wenn man gesprochene Sprache analysieren und verstehen möchte
- Schwieriges Problem
- Oft führen Umgebungsgeräusche oder ungünstige akustische Bedingungen zu Audiosignalen, die durch Rauschen und Störeinflüsse schwierig zu decodieren sind

- Ähnliche Effekte kennen wir beim Menschen
- Menschen benutzen dabei zusätzliche Informationen wie Lippenbewegungen (und eben Gesten)
- Auch maschinelle Spracherkennung kann zusätzliche (Kontext-) Informationen ausnutzen
 - Kamera filmt den Sprecher

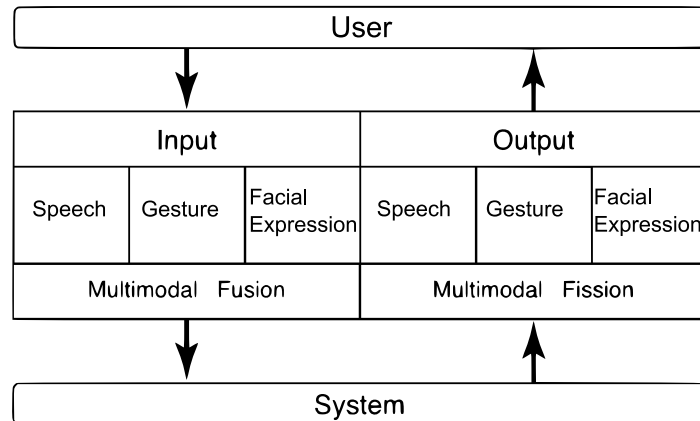
Multimodale Systemausgaben

- Ausgaben geprägt durch visuelle und akustische Medien
- Vereinzelt haptisches Feedback
- Soundausgabe:
 - Geräusche
 - Musik
 - Sprache
 - Kurze Audiosequenzen (Earcons)
- Visuelle Medien in der Regel reichhaltiger
 - Photos
 - Graphiken
 - Texte
 - Animationen
 - Icons
- Haptisches Feedback
 - Spiele mit Force Feedback
 - Bei Einschränkung der Sehfähigkeit
 - Mobile Geräte

Kombination

- Multimodale Ausgabe einfacher zu realisieren als Eingabe
- Im Prinzip sind einfache Kombinationen bereits multimodal bzw. -codal in unserem Sinne
 - Icons mit Text (Multicodal)
 - Videos mit Bild und Ton (Multimodal)
- Bei Spielen werden oft Graphik und Sound gemischt, um das Geschehen zu untermalen
 - Realistischer
 - Einfachere Interaktion
 - Erleichterung der Immersion
- Multimodale Systeme im engeren Sinne sind solche, bei denen die multimodale Präsentation online, also bei Bedarf, neu generiert wird
- Entsprechendes gilt für die Eingabe (Online-Kombination verschiedener Modalitäten)

Multimodale Systeme



Multi-modal interaction in SmartKom, according to [Malaka et al. \[2009\]](#)

Virtuelle Personen

- Beispiel SmartKom: Nutzer und System können die gleichen Modalitäten nutzen
- Man kann mit dem System sprechen und gleichzeitig wird Gestik und Mimik analysiert
- Das System kann durch virtuelle Repräsentationen selber verschiedene Modalitäten benutzen
- Prototyp für unterschiedliche Szenarien
 - Fußgänger mit Smartphone
 - Informationskioske
 - Unterhaltungselektronik

Symmetrische Multimodale Kommunikation



Abbildung 9.3: Interaktionsagent von SmartKom (linkes Bild). Dieser Agent kann selbst Gesten und Mimik darstellen und spricht mit dem Nutzer. Bild Mitte und rechts stellen zwei Systemvarianten von SmartKom dar, einmal für Fußgänger und einmal an einem öffentlichen Informationskiosk.

Roboter

- Beispiel Autonomous City Explorer: Ein Roboter, der seinen Weg selber findet
- Der Roboter weiß nur, wo er hin soll, hat aber keine Karte und kein GPS
- Er bittet unterwegs um Hilfe, fragt nach dem Weg
- Man kann dem Roboter den Weg zeigen

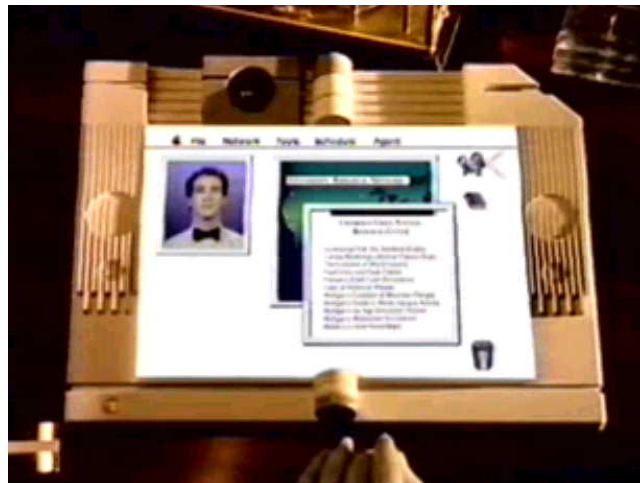
Video 14.2: Multimodale Roboter



Autonomous City Explorer (04:30)

2 Natürliche Sprache

Video 14.3: Apple Knowledge Navigator



<http://v.gd/knowledgenavigator> (5:45)

Video 14.3: Apple Knowledge Navigator

- First described in 1987 by Apple CEO at the time Jon Sculley in his book Odyssey
- Inspired by works of Vannevar Bush and Alan Kay
- Modern communication technology, connecting computers and databases worldwide giving more user more information and wider information channels
- Real-time 3D animations for visualizing complex models
- Better database technologies as key to copious information systems
- Hypermedia, connecting text, images, sound, and video, helping future users to navigate vast information spaces
- Artificial Intelligence as key component
- AI makes it possible to use agents to recognize personal preferences of users and propose problem solving strategies, leading to an increase in productivity

Natürliche Sprache

- Natürliche Sprache das Kommunikationsmedium, das Menschen am häufigsten benutzen, wenn sie mit anderen Menschen Informationen austauschen
- Gesprochen oder geschrieben
- Herausforderung: nicht nur digitale Repräsentation, sondern “verstehen”
- Menschliche Sprache enthält nicht nur explizite Informationen, die eindeutig durch Worte festgelegt sind
- Bedeutung liegt oft im Kontext
- Zudem spielen Sprachmelodie, Gestik, Mimik eine Rolle
- Derzeit noch hohe Fehlerraten
 - Allerdings: Systeme wie Apples Siri oder Google Voice Search für eingeschränkte Kontexte eingeführt
 - Ebenso automatische Sprachsteuerung in Call Centern
- Geschriebene Sprache z.B. bei Suchmaschinen erfolgreich
 - Im Regelfall ergänzt durch formale Operatoren
 - Auch hier werden Kontextinformationen herangezogen

Gesprochene Sprache

- Bei gesprochener Sprache werden in Mund und Rachenraum Klänge und Geräusch gebildet
- Diese werden als Sprache wahrgenommen
- Nicht völlig beliebige Geräuschsequenzen
- Sprache besteht aus einer begrenzten Menge an Phonemen
- Im deutschen und englischen etwa 40
 - Vokalphoneme
 - Konsonantenphoneme
- Einheitlich für alle Sprecher, aber mit individuellen Unterschieden
 - Alter, Dialekt, Geschlecht
- Darüber hinaus sind Sätze oft nicht wohlgeformt
 - Korrekturen oder Füller (“äh”)

Geschriebene Sprache

- Im Gegensatz zu gesprochener Sprache basiert geschriebene Sprache auf einem Alphabet, das nicht unbedingt identisch zu den Phonemen ist
- Handschrift hat ebenso wie gesprochene Sprache erhebliche individuelle Unterschiede
 - Rechtschreibfehler und unterschiedliche Schreibweisen
- Gelingt die Analyse und Erkennung von Phonemen und Buchstaben, dann werden nach morpho-syntaktischen Regeln Wörter gebildet
- Diese formen nach syntaktischen Regeln Sätze
- Ein Satz kann interpretiert werden und ihm eine Bedeutung zugewiesen werden
- Pragmatik und Kontext können ebenfalls eine Rolle spielen

Erkennung und automatische Verarbeitung

- Ziel ist häufig die Übersetzung in eine logische Form zur Beschreibung von Sachverhalten
- Menschliche Sprache hat in der Regel mehr Ausdrucksmöglichkeiten als logische Kalküle
 - Prosodie, Wahl des Ausdrucks
- Daher gehen bei der Transformation Informationen verloren
- Besonderes Problem sind Mehrdeutigkeiten (Ambiguitäten)
 - Mehrdeutigkeit: “Bauernopfer”
 - Ellipsen: “Ruhe jetzt”
 - Metaphern: “Der Bundestag debattiert”

NLP

- NLP – Natural Language Processing
- Typische Systeme zur Verarbeitung natürlicher Sprache orientieren sich an klassischen semiotischen Ebenen
 - Zeichen/Phoneme
 - Morpho-Syntax
 - Semantik
 - Pragmatik
 - Kontext
- Auf den verschiedenen Ebenen werden unterschiedliche Verfahren und Datenstrukturen verwendet
- Die Computerlinguistik beschäftigt sich mit der Analyse und Generierung von Sprache mit algorithmischen Methoden
- Die im folgenden vorgestellte Pipeline ist vereinfacht, teilweise sind weitere Schritte notwendig

Phonemerkennung

- Analoges Sprachsignal durch ein Aufnahmegerät und Digitalisierungsverfahren in ein digitales Signal umwandeln
- Überführung in ein Spektrogramm, da dieses charakteristisch für die unterschiedlichen Phoneme ist
- Überführung des zeitabhängigen Audiosignals in eine diskrete Symbolfolge
- Suche nach Phonemen
- Einsatz von Mustererkennern, die auf vielen Beispieldaten beruhen
- Diese kann sein:
 - Sprecherabhängig
 - Sprecherunabhängig (wesentlich schwieriger)

Audiosignal

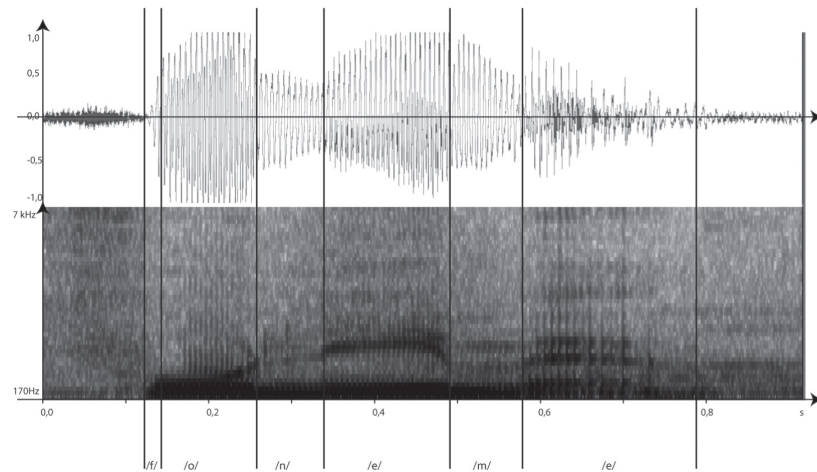
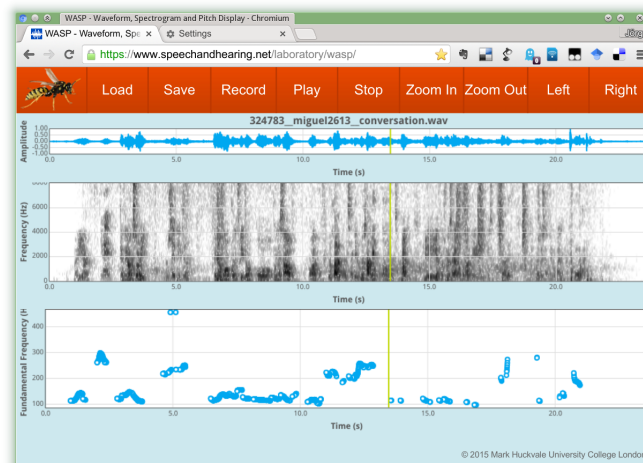


Abbildung 9.4: Wellenform und Frequenzspektrum des gesprochenen Wortes „Phoneme“. Dabei sind die einzelnen Phoneme des Wortes durch vertikale Linien abgegrenzt.

Analyse-Werkzeug



www.speechandhearing.net/laboratory/wasp

Worterkennung

- Untersuchung auf mögliche Wörter
- Typischerweise statistische Verfahren: wie wahrscheinlich ist der Übergang von einer gegebenen Folge von Phonemen zu einem neuen Phonem
 - z.B. Hidden Markov Models
- Wegen Ungenauigkeiten in der Verarbeitungskette wird mit Wahrscheinlichkeiten gearbeitet
- Nicht unbedingt eine eindeutige Liste von Worten, sondern Hypothesen für Wörter (Worthypothesen-graph)
- Modelle für alle Wörter des Vokabulars
 - Einfache Systeme beschränken sich auf Kommandowörter
- Für einfache Systeme reicht diese Verarbeitungskette aus
- Komplexere Systeme betrachten weitere Merkmale

Satzerkennung

- Bei der Satzanalyse werden teilweise symbolische Verfahren eingesetzt
- Auf Grundlage von Worthypothesen sollen die Bedeutungen erkannt werden
- Dabei können z.B. formale Grammatiken (Chomsky) eingesetzt
- Neben den bekannten Grammatiken aus der Chomsky-Hierarchie werden in der Computerlinguistik eine ganze Reihe unterschiedlicher Grammatiken und Formalismen eingesetzt
 - Parser für die Erkennung von Sätzen
- Andere Ansätze benutzen graphentheoretische oder statistische Ansätze (Hidden Markov Models, Bayesian Networks)
- Statt Sätzen werden im allgemeinen bedeutungstragende Einheiten erkannt, die auch über Satzgrenzen gehen können oder Sätze teilweise aufteilen können

Bedeutungserkennung

- Identifikation der Bedeutung der erkannten Sätze und Satzteile
 - “Siri, call Steve!” – Welche Funktion soll ausgelöst werden, welcher Kontakt soll angerufen werden
 - “OK Google, how tall is Barack Obama?” – Identifikation der Intention der Suche (Körpergröße), Identifikation der Person, über die Informationen gewünscht werden
 - “Alexa, do I need an umbrella today?” – Identifikation der Intention (Regenschirm bei Regen, ergo Nachfrage nach dem Wetter notwendig)
- Sowohl statistische Verfahren als auch explizite semantische Modellierung
 - Ontologien
 - Wissensgraphen

Weitere Elemente

- Die skizzierte Verarbeitungskette aus
 - Phonemerkennung
 - Worterkennung
 - Satzerkennung
 - Bedeutungserkennungist stark verkürzt
- In der Praxis notwendige Teilschritte können umfassen
 - *Tokenisierung* – Verallgemeinerung der Worterkennung, so können z.B. lateinische Namen von Arten bestehend aus Genus und Species ein Token sein
 - *Part-of-speech-tagging*, eliminieren von Füllwörtern und Erkennung sinntragender Einheiten, welche nicht unbedingt Satzgrenzen folgt
 - *Stemming*, das ist die Rückführung von Worten auf eine Grundform
 - *Entity recognition*, das ist die Erkennung von Objekten und Personen, z.B. mit Hilfe von Ontologien

Sprachausgabe

- Im Prinzip der umgekehrte Weg
- Unterscheidung in zwei Schritte
 1. **Sprachgenerierung**: von der Intention zu Wortketten
 2. **Sprachsynthese**: von der Wortkette zum Audiosignal
- Scheint leichter zu sein
- Modelle wie bei der Erkennung, z.B. formale Grammatiken, graphentheoretische Modelle, Frames, Scripts

- Sprachsynthese mittels unterschiedlicher Verfahren
 - Aufgenommene Samples von Wörtern oder Satzfragmenten (Bahnhofsdurchsagen)
 - Komplette synthetische Erzeugung (viele Screenreader)
- Problem beim Einsatz aufgenommener Sprache ist häufig die fehlende Flexibilität
- Problem der Synthese ist oft der künstliche Eindruck (Roboterstimme)

Automatische Sprachverarbeitung

- Einsatz von NLP in ganz unterschiedlichen Anwendungen kann möglich, sinnvoll oder gar notwendig sein
- Besonders sinnvoll in Situationen, in denen andere Interfaces schlecht zu nutzen sind
 - Mobile Geräte, wenn Maus und Tastatur nicht nutzbar sind
 - Wenn die Hände beschäftigt sind
 - Wenn der Blick auf andere Dinge gelenkt werden muß
 - Wenn andere E/A-Geräte nicht zur Verfügung stehen
- Neben situationsbedingten Gründen auch individuelle
 - Sehbehinderte Nutzer
 - Einschränkungen physischer Möglichkeiten
 - Unkenntnis der Bedienung von Systemen
- Weiterhin: Aufgabenspezifische Gründe (inhaltsbasierte Suche, freie Dialoge)

Consumer-Bereich

- Sprechende Akteure in Computerspielen sind unterhaltsamer und lebendiger
- Teilweise auch Einsatz bei der Steuerung von Geräten der Haushaltselektronik
 - Sprachsteuerung von Autoradios
 - Sprachsteuerung von Smartphones
 - Amazon Echo und vergleichbare Geräte
 - * Aktuelle Nachrichten, Wetterbericht
 - * Steuerung Mediensystem
 - * Bestellungen
- Ansonsten sind Sprachsteuerungen im privaten Haushalt eher noch die Ausnahme
 - Geringe Akzeptanz bei hoher Fehlerrate
 - Sinnvolle Anwendungsbereiche – ist es wirklich notwendig, mit dem Fernseher oder der Kaffeemaschine zu sprechen?

Vor- und Nachteile

- Vorteile
 - Leicht zu erlernen und zu merken
 - Sprache sehr ausdrucksmächtig
 - Oft schnell und effizient (aber nicht immer)
 - Kein oder nur kleiner Bildschirm benötigt
- Nachteile
 - Erkennung und Interpretation oft noch nicht zuverlässig
 - Wissen über den Diskursbereich wird beim Nutzer vorausgesetzt
 - Geschriebene Sprache benötigt eine Tastatur
 - Erweiterungen und neue Funktionalitäten sind nicht sichtbar (discoverability)
 - Teuer in der Implementierung
- Bei schlechter Umsetzung können die Nachteile die Vorteile schnell aufheben, das System wird zum Mißerfolg

Abhilfe

- Verbesserungen der Robustheit
 - Einschränkungen beim Vokabular
 - * Es ist einfacher, eine kleine Menge an Kommandos zu erkennen
 - Dynamischer Austausch von Grammatiken und Lexika je nach Kontext
 - Strategien des Interaktionsmanagements, z.B. geschickte Frageformulierung
- Fehlerbehebungsstrategien
 - Anpassung an möglicherweise auftretende Probleme
 - * Beantwortung einer Frage dauert länger, das sollte dem Benutzer gezeigt werden
 - Benutzertests
 - * Vor der Implementierung z.B. durch Wizard of Oz-Experimente
 - * Das System wird von menschlichen Testern simuliert
 - * Nebeneffekt: Daten werden gesammelt

3 Virtuelle Welten

Video 14.4: Microsoft HoloDesk



Microsoft HoloDesk – Direct 3D Interactions (4:14)

Video 14.5: SixthSense



Pattie Maes + Pranav Mistry: Meet the SixthSense interaction – TED (8:36)

Virtual Reality

- Bereits kennengelernt: Begriff der Immersion
- Abhängig u.a. von:
 - Wie viele Sinne werden angesprochen?
 - Wie realistisch ist die Darstellung?
- Bei Filmen mit guter Ton- und Bildqualität gelingt die Immersion häufig schon sehr gut
- Das Medium bleibt aber passiv
- Positionen durch Kamera festgelegt, man kann sich nicht selber “umsehen”
- Systeme der Virtuellen Realität (VR) haben den Fokus auf der visuellen Darstellung und dem visuellen Eintauchen in eine virtuelle Szene
- Häufig verwendet:
 - Head-Mounted Displays
 - Cave

HMD und Cave

Head-Mounted Display: Kleine Displays, die vor den Augen getragen werden, in Form einer Brille. Durch die Nähe zum Auge kann ein großer Sichtbereich simuliert werden, die Auflösung ist aber begrenzt. Mit zwei Displays sind Stereobilder möglich. Einige Modelle erlauben den Blick auf die Umwelt (durch das Display oder daran vorbei), andere decken vollständig ab.

Cave: Eine Cave ist eine Installation, in der große Displays um den Nutzer herum angeordnet sind. Typisch: Kubus mit 3 (rechts, links, vorne) bis 6 (um den Nutzer herum) großen Projektionsflächen. 3D-Darstellung durch Shutterbrillen oder Polarisation.

HMD: 1



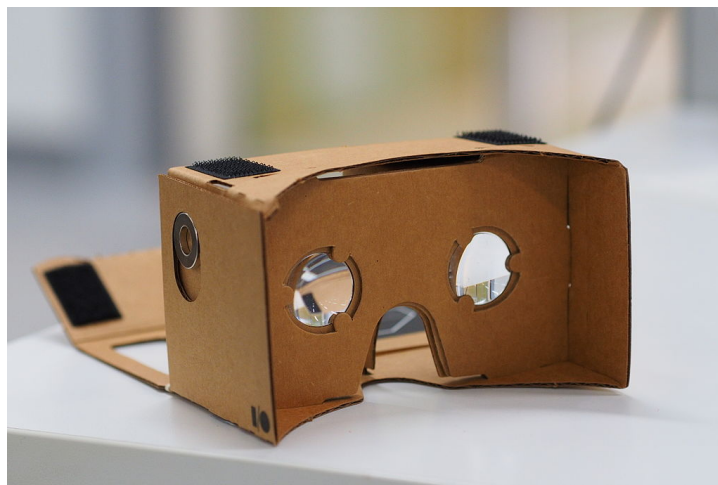
Abbildung 9.7: Beispiel eines Head-Mounted Displays (Foto: TZI)

HMD: 2



 Popular Mechanics

HMD: 3



 Wikipedia

HMD: 4



 Wikipedia

HMD: 5



Foto: Zeiss

HMD: 6



Foto: Microsoft

Video 14.6: HMD



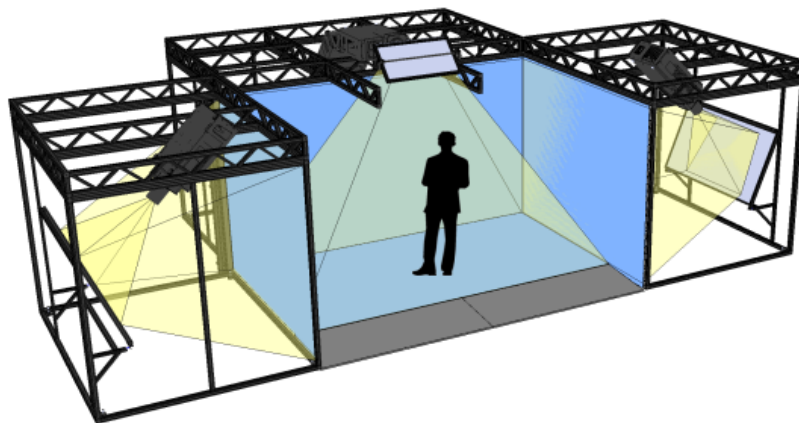
VR HMD – Computerphile (11:34)

Cave: 1



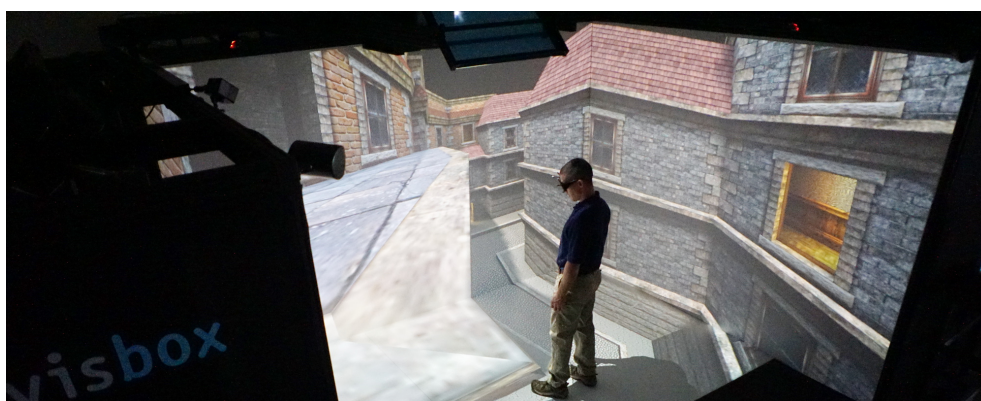
Abbildung 9.8: Nutzer einer Cave mit einem visuellen 3D-Tracking für Handgesten
(Foto: im.ve, Universität Hamburg)

Cave: 2



Visbox: Schemazeichnung

Cave: 3



Visbox: Beispiel

Video 14.7: CAVE



VR CAVE – Computerphile (11:34)

Tracking

- Bei einer Cave wichtig: Position des Kopfes genau treffen
 - Für einen einzelnen Nutzer optimiertes (Stereo-) Bild
 - Auch für Gestensteuerung verwendbar
 - Spezielle 3D-Mäuse
 - Einsatz vor allem in speziellen Anwendungskontexten (Exploration von Ölfeldern, Architektur)
- Bei den HMD ist es wichtig, neben der reinen Kopfposition andere Elemente zu erkennen (Neigung und Blickrichtung)
 - Oculus mit der Rift HTC mit der Vive haben dieses Gebiet wiederbelebt und für Privatanwender interessant gemacht
- Alle Medien und Modalitäten nutzbar, die sinnvoll sind
- Im Prinzip kann auf visuelle Eindrücke verzichtet werden (Audio-VR)
- Allgemein soll hohe Immersion angestrebt werden
 - Komplexe Systeme, die Video, Audio und Haptik vereinen
 - Einfachere Systeme, z.B. für PC-Spiele

Augmented Reality

- Im Gegensatz zur Virtuellen Realität kann man auch die natürliche Immersion in die reale Umwelt zur Basis nehmen und diese anreichern
 - Teile der virtuellen Welt in die reale überführen
 - Wenn die Sinne (in der Regel das optische System) direkt angesprochen werden spricht man von erweiterter (augmentierter) Realität
- Beispiel
 - Teildurchlässige Brille
 - Rekonstruktionen historischer Gebäude/Artefakte an den Stellen, an denen diese sich vorher befunden haben
- Google Glass und Microsoft HoloLens bekannteste Systeme, welche für Konsumenten nutzbar sind
- Alternativ zu Systemen, die mit HMD arbeiten, ist auch die Nutzung von Projektionen möglich

Beispiel: Architektur

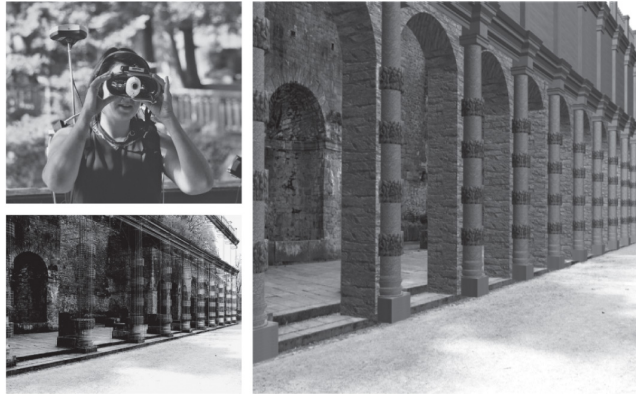


Abbildung 9.9: Beispiel für Augmented Reality. Nutzer erhalten eine Brille, durch die einerseits die Umwelt gesehen werden kann, in die aber auch Überblendungen von computergenerierten Inhalten eingeblendet werden können. In GEIST werden so zu Ruinen von Gebäuden Rekonstruktionen von Gebäudeteilen eingeblendet.

Beispiel: Ingress



📱 Niantic. Tip: Be Enlightened.

Beispiel: Pokémon GO



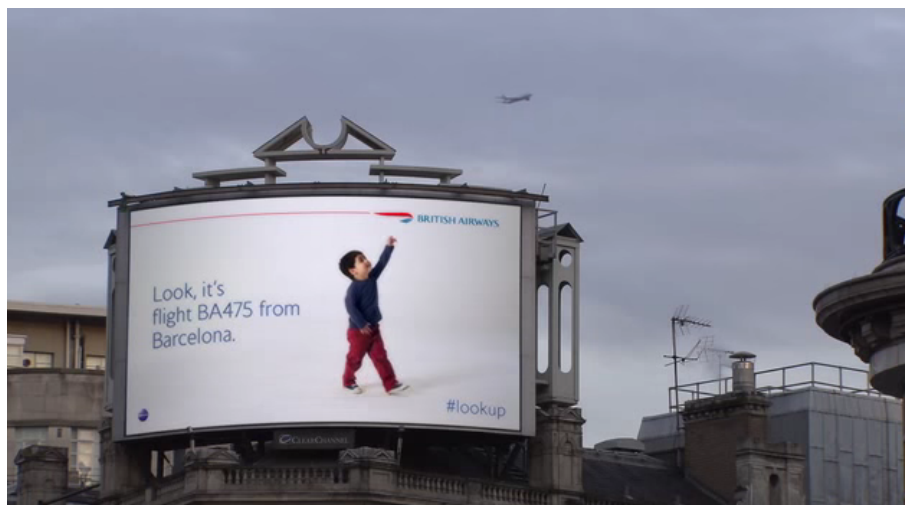
pokemongo.com

Video 14.8: Microsoft HoloLens Project X-Ray



[Microsoft HoloLens Project X-Ray \(4:02\)](#)

Video 14.9: Look Up



[British Airways: lookup in Piccadilly Circus \(0:26\)](#)

Augmented Virtuality

- Umgekehrt kann man die Virtuelle Realität zur Basis nehmen und diese mit Objekten aus der natürlichen Welt anreichern
- Beispiel
 - HTC Vive unterstützt physische Gamecontroller, die in der virtuellen Realität eingeblendet werden
- Stichwort Mixed Reality:
 - Man kann von einem Kontinuum sprechen – von der Natürlichen Umwelt bis zu vollständig virtuellen Realitäten

Mixed Reality

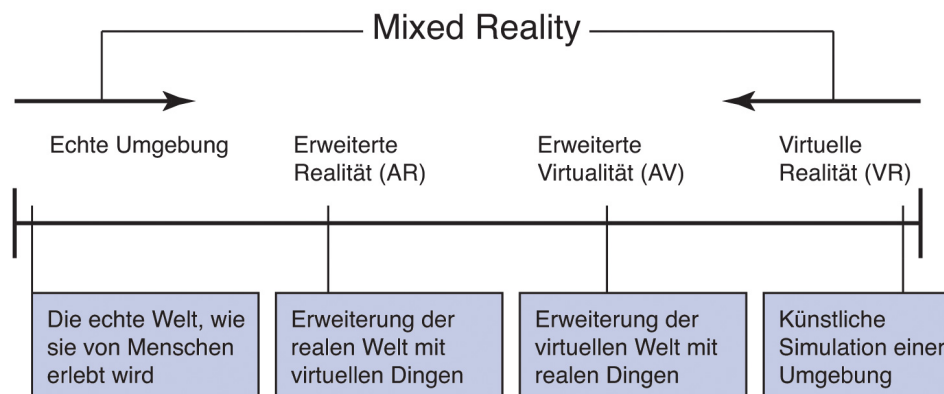


Abbildung 9.10: Kontinuum der Mixed-Reality-Paradigmen nach Milgram und Kishino. Auf den Endpunkten sind die reale Welt ohne digitale Zusätze und die virtuelle Welt, die komplett simuliert wird.

Einsatzgebiete

- VR-Simulation bei der Entwicklung neuer Produkte, z.B. Autos oder komplexe Maschinenanlagen
- VR-Trainingssimulationen, z.B. bei Piloten
- VR-Spiele für Einzelnutzer oder große Gruppen, bei denen in simulierten Welten mit animierten virtuellen Charakteren gespielt wird
- Spiele, die reale oder real anmutende Gegenstände als EA-Geräte nutzen
- Sind Navigationssysteme bereits AR-Systeme?
- Frage: wie können realistische Erlebnisse von Nutzern digitaler Medien durch die Einbindung realer und immersiver virtueller Komponenten wirkungsvoll und nachvollziehbar gestaltet werden?

Video 14.10: Cyberith Virtualier

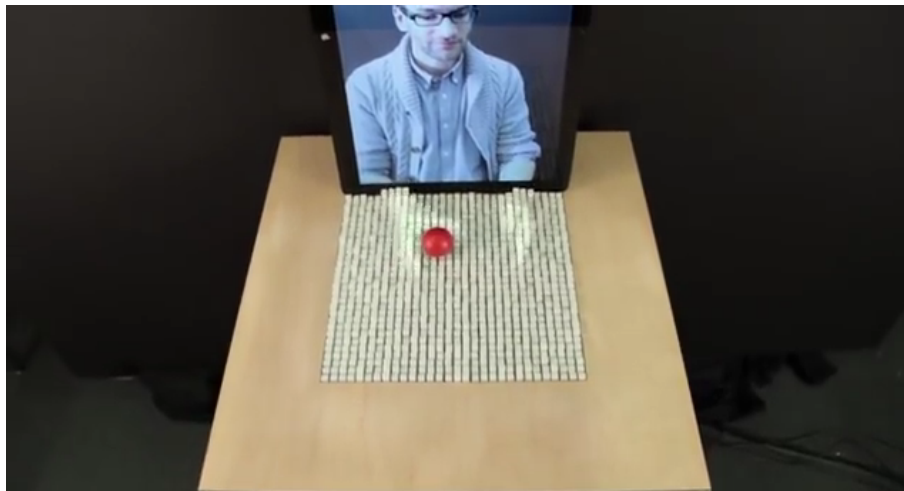


Cyberith Virtualizer oder: Mein Zimmer ist zu klein (2:00)

Weitere Ansätze

- “Bridging the Digital-Physical Divide”
- Tangibles und Ambiente Systeme bringen ebenfalls virtuelle und reale Objekte zusammen
- Wie bei AR-Systemen ist die reale Immersion die Basis
- Fokus liegt hier eher darauf, daß reale Objekte mit ihren physischen Eigenschaften genutzt werden, um mit digitalen Objekten zu interagieren
- Einerseits: Immer mehr “virtuelle” Anreicherungen der realen Welt
- Andererseits: Einfügen realer Objekte in virtuelle Welten bringen diese hinein in die virtuelle Welt

Video 14.11: inFORM Dynamic Shape Display



inFORM – Interacting With a Dynamic Shape Display (3:41)

4 Ambient Computing

Video 14.12: Microsoft Productivity



Microsoft: Productivity Future Vision (2011) (6:18)

Video 14.13: Microsoft Manufacturing



Microsoft: Manufacturing Future Vision (3:26)

Ambiente Medien

- Digitale Medien entwickeln sich zu ubiquitären Medien, die immer und überall verfügbar sind
- Klassische Computersysteme an Tastatur, Maus und Bildschirm gebunden
- Heute: Verfügbarkeit auf vielen unterschiedlichen Geräten, die teilweise in den Hintergrund treten
- Dieser Umbruch wird auch als **Ubiquitous** oder **Ambient Computing** bezeichnet
- Pro Person stehen mehrere und vernetzte Geräte zur Verfügung
- Computer nicht mehr unbedingt als solche erkennbar oder eingebettet in andere Systeme

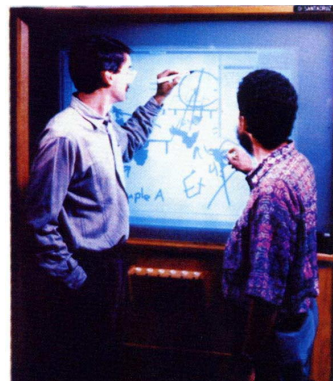
Ubiquitous Computing

“The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.”

(Weiser [1991], *The Computer for the 21st Century*)

“This then is Phase I of ubiquitous computing: to construct, deploy, and learn from a computing environment consisting of tabs, pads, and boards. This is only Phase I, because it is unlikely to achieve optimal invisibility.”

(Weiser [1993], *Some Computer Science Issues in Ubiquitous Computing*)



Weitere Begriffe

- **Ambient Intelligence:** “Europäischer” Begriff, der im Verhältnis zu Ubiquitous Computing die Nutzung der KI stärker betont
- **Pervasive Computing:** Computer in Alltagsobjekten (durchdringend); häufig auf Aspekte der IT-Architektur abhebend
- **Smart Spaces:** Intelligente, aber räumlich abgegrenzte Gebiete (nicht ubiquitär)
- **Mobile Computing:** Schwerpunkt liegt auf der Mobilität der Nutzer oder der Geräte
- **Wearable Computing:** Computer als Teil der Kleidung
- **Kontext:** Einbeziehung der Situation, in der sich das System oder der Nutzer befindet

Working with Context

Context Awareness

Trying to **detect** the situation the system is in.

- **Example:** An ambient intelligent system for supporting health personnel figures out that the user is on a ward-round because of the time of the day, the location, and the other persons present.

Context Sensitivity

Acting according to the situation the system thinks it is in.

- **Example:** the same system fetches the newest versions of electronic patient records of all patients in the room from the hospital systems. When the user stands close to the bed of a patient, the system automatically displays them.

Best Practice Context Models

- **Environmental context:** This part captures the users surroundings, such as things, services, people, and information accessed by the user.
- **Personal context:** This part describes the mental and physical information about the user, such as mood, expertise and disabilities.
- **Social context:** This describes the social aspects of the user, such as information about the different roles a user can assume.
- **Task context:** the task context describe what the user is doing, it can describe the user's goals, tasks and activities.
- **Spatio-temporal context:** This type of context is concerned with attributes like: time, location and the community present.

Ambient Intelligent Systems

At the core of an ambient intelligent system lies the ability to **appreciate the system's environment**, be **aware of persons** in this environment, and **respond intelligently to their needs** (Ducatel et al. [2001], *ISTAG Scenarios for Aml in 2010*).

- **Perception:** The initial act of perceiving the world that the system inhabits
- **Context Awareness:** Being aware of the environment and reasoning about ongoing situations
- **Context Sensitivity:** Exhibit appropriate behaviour in ongoing situations
- **Action:** Changing the environment according to context

Ambient and Other Systems

- Ambient Systems will not replace, but augment existing systems and their interaction capabilities
- Examples:
 - Location-based services can enhance the usefulness of existing systems
 - An ambient system might “fall out of ambience” if something goes wrong



Beispiel: Intelligente Häuser

- Intelligente Wohnumgebungen “hot topic” in der Forschung
- Viele Projekte im Bereich “Ambient Assisted Living”
 - Demographischer Wandel, “Überalterung der Gesellschaft”
 - Heutiger Pflegeaufwand zu kostspielig
 - Menschen längeres Verweilen in der gewohnten Umgebung ermöglichen
- Räume und Häuser mit einer Reihe von Sensoren ausgestattet
 - Wer ist in einem Raum
 - Was macht die Person
- Kontextualisierte Änderung
 - Licht, nur wenn jemand da ist
 - Musik, die mitgeht
 - Abfrage/Steuerung von Haushaltselektronik

Beispiel: EMBASSI



- Projekt EMBASSI
- Im Beispiel sagt eine Person “heller bitte”
 - Zeigt auf Fernseher
 - Zeigt auf Licht

Beispiel: Stadt

- Weiteres beliebtes Anwendungsgebiet
- Ubiquitäre Unterstützung von mobilen Nutzern in einer Stadt
 - Navigationshinweise
 - Tips für Stadtbesichtigungen
 - Hinweise auf Einkaufsmöglichkeiten bzw. Restaurants
 - Informationen anderer Nutzer (mobile Blogs, virtual tagging)
 - Nachrichten zu Themen, die den Nutzer interessieren
- Da Displays mobiler Geräte klein sind wird daran gearbeitet, Nutzern Zugriff auf stationäre Displays zu geben
 - Public Displays
 - Cross-Device Interaction (XDI)
- Objekte können über elektronische Tags (z.B. RFID, Bluetooth-Beacon) auf weitere Information über sie verweisen

Video 14.14: A Day Made of Glass



Corning: A Day Made of Glass 2 – Unpacked (11:25)

5 Intelligente Medien

Intelligente Medien

What is Artificial Intelligence (AI)?

“It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.” (McCarthy)

- Methoden der Künstlichen Intelligenz können für digitale Medien vielfach verwendet werden
 - Hilfsmittel bei der Erstellung (Authoring Tools)
 - Intelligente Benutzungsschnittstellen
 - Personalisierte und kontextualisierte Dienste
 - Archivierung, Verwaltung, Suche in digitalen Medien
- Verwendete Techniken u.a.
 - Wissensmodellierung und Schließen
 - Maschinelles Lernen
 - Planung, Suche, Analyse, Klassifikation

Authoring

- Beispiele für Werkzeuge, die auf KI-Techniken beruhen
 - Smarte Auswahlwerkzeuge: Um in Bildern oder Filmen Personen oder Objekte freizustellen können Verfahren zur Unterscheidung von Vorder- und Hintergrund eingesetzt werden
 - Sprach- und Gestenerkennung
- Es gibt auch Ansätze, den kompletten Prozeß der Erstellung zu automatisieren
 - Rohdaten wie Bilder, Texte, Graphiken, modellierte Objekte werden analysiert und für die Darstellung aufbereitet
 - Typischerweise muß hier auch modelliert werden, wie man abhängig von kommunikativen Zielen Inhalte in einem Medium repräsentiert
 - Bei Einsatz von KI-Systemen können die einzelnen Komponenten wesentlich flexibler und eigenständiger erzeugt werden (Anpassung an unterschiedliche Geräte, Zusammenfassungen, Nachrichtenübersichten)

Ablauf

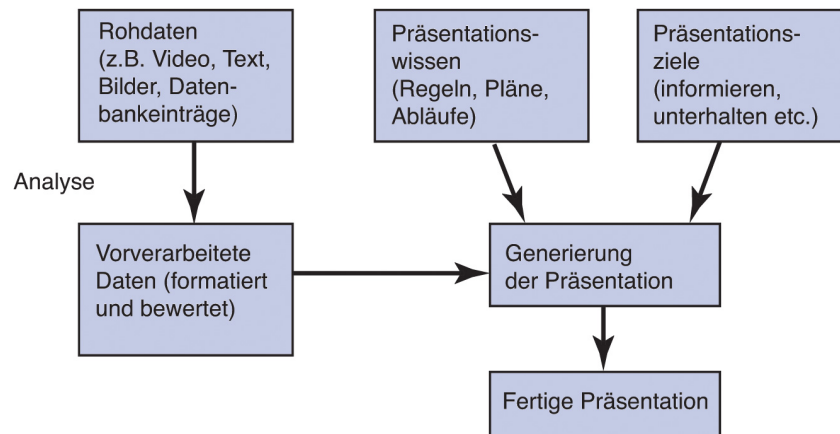


Abbildung 9.11: Prototypischer Ablauf einer automatischen Präsentationsgenerierung

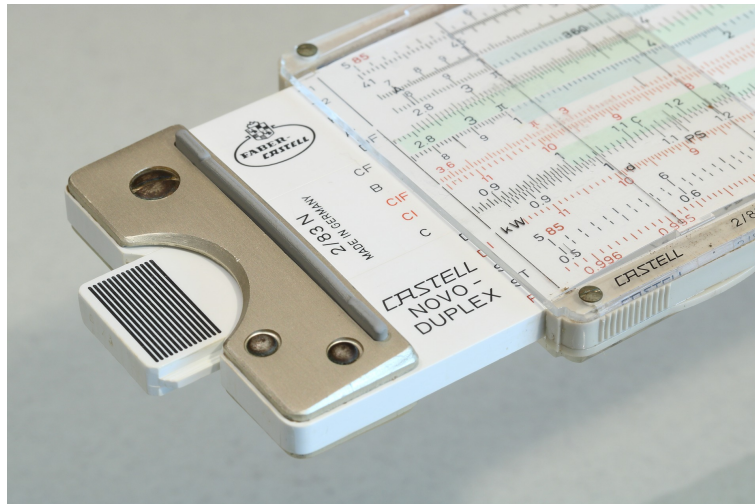
Beispiel: Generierung von Filmtrailern

- Rohdaten: kompletter Spielfilm inklusive Audio- und Videospuren sowie Untertitel, darüber hinaus Zitate und Namen der Schauspieler und Regisseure von IMDB
- Automatische Erkennung von Szenen und Einstellungen
 - Kameraeinstellung
 - Dialog, Action?
 - Szenen mit Schlüsselsätzen
- Präsentationswissen, welches die Struktur der Trailer von Actionfilmen modelliert, inkl. der akustischen und visuellen Effekte
- Kommunikatives Ziel ist die Trailergenerierung
- Zunächst: Plan für den Trailer, dann Suche nach passenden Einstellungen, Erzeugung von Effekten
- Bei Tests mit menschlichen Nutzern wurden die automatisch generierten Trailer nicht schlechter bewertet als von Menschen erstellte

Intelligente Interaktion

- Intelligente Benutzungsschnittstellen sollen Anpassung, Verständnis und Vermittlung von Benutzungsschnittstellen verbessern
- Zwei Möglichkeiten
 - System so intelligent wie möglich zu machen, damit es auf Nutzeranfragen möglichst gute Lösungen präsentieren kann
 - * Analogie: Taschenrechner mit logarithmischen Funktionen
 - Interaktionsform intelligent gestalten, so daß der Nutzer das Problem selber gut lösen kann
 - * Analogie: Rechenschieber
- Bei der ersten Variante ist die Intelligenz im System, während in der zweiten die Problemlösekompetenz weitgehend beim Nutzer verbleibt

Rechenschieber



Rechenschieber (via olafpictures)

Smarte Interaktion: Ein Kontinuum

- Grob kann man unterschiedliche Varianten der Interaktion unterscheiden
 - Klassische konventionelle Systeme: Alle Intelligenz bleibt beim Nutzer
 - Klassische KI: Die Intelligenz wandert komplett in das System
 - Intelligente Benutzungsschnittstellen: Problemlösekompetenz beim Benutzer, klassische Berechnung beim System, "Intelligenz" in der Schnittstelle
 - Intelligente Kooperative Systeme: Teile der Problemlösekompetenz jeweils bei System und Nutzer mit intelligenter Schnittstelle – "human in the loop"
- Es geht letztlich immer darum, die Kompetenz, Effektivität und Effizienz des gesamten sozio-ökonomischen Systems zu verbessern
- (bei sicherheitskritischen Systemen) auch ethische Frage: wer hat die Verantwortung bei autonomen Fahrzeugen, sollten wir den "human in the loop" sicher stellen?

Varianten

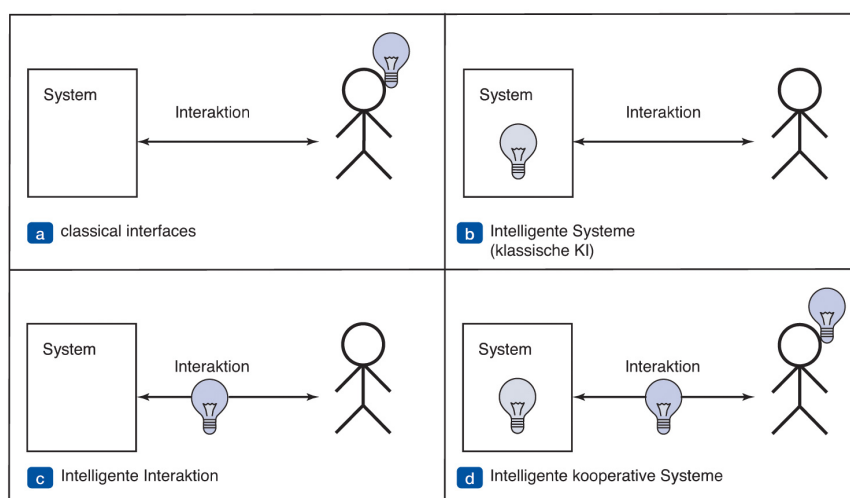


Abbildung 9.12: Verschiedene Sichten auf intelligente Interaktion

Intelligente Dienste

- Nutzeranpassung: Modelle oder Profile der Nutzer
 - Von Hand oder automatisch erstellt
- Kontextanpassung
 - Navigationssystem im Auto, im Zug oder zu Fuß nutzen?
- Integration von Diensten
 - Intelligenter Routenplaner, der Verkehrsinformationen, Wetterdaten, Fahrpläne öffentlicher Verkehrsmittel so integriert, daß Reisen mit unterschiedlichen Verkehrsmitteln vorgeschlagen werden
- Medienretrieval
 - Für die Suche nach Filmszenen reicht eine einfache Suche nach Schlüsselwörtern nicht aus

Video 14.15: Intelligente Bilderzeugung

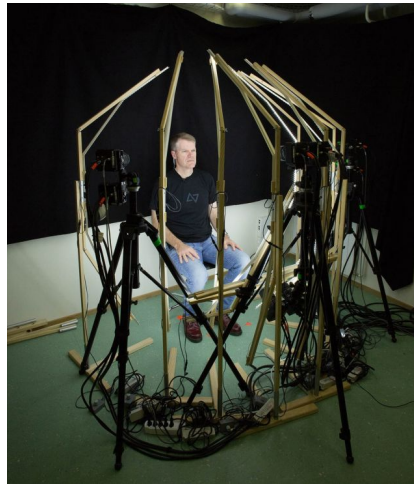


Design FX/WIRED: How 'Rogue One' Recreated Grand Moff Tarkin (6:25)

ILM Face Pipeline

1. Digitalisierung – Digitales Modell eines Schauspielers
 - “Lightning Rig” – Diffuse und Glanzreflexion der Haut
 - “Medusa Rig” – Gesichtsausdrücke und deren Übergänge
2. Motion Capture – animieren digitales Modell
 - Mehrere Kameras am Kopf nehmen Gesichtsausdrücke auf
 - I.d.R.: Digitales Modell von z.B. Warcraft-Charakteren
 - Kann auch Modell eines anderen Schauspielers sein
3. Snap Solver – Automatisches Nacharbeiten
 - Kleine Änderungen – Minimiert Abweichungen des digitalen Modells von den Mocap-Daten (Gesichtsausdrücke)
 - Große Änderungen – Künstliche Charaktere (z.B. Warcraft) weitgehend automatisch animieren
4. Manuelle Animation – Nacharbeiten
 - Notwendig wenn digitales Modell und Motion Capture von verschiedenen Schauspielern stammen
 - Ausgleichen von Unterschieden, z.B. welche Muskeln genutzt werden, Phoneme auszudrücken

Face Pipeline: Digitalisierung



Medusa Rig ([fxguide.com](https://www.fxguide.com))

Face Pipeline: Motion Capture



Motion Capture Filming ([fxguide.com](https://www.fxguide.com))

Face Pipeline: Digitales Modell



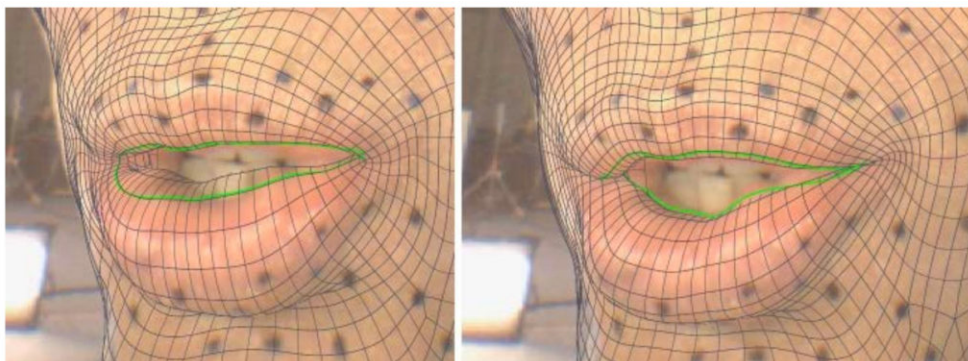
Generate Digital Model ([fxguide.com](https://www.fxguide.com))

Face Pipeline: Snap Solver "Grob"



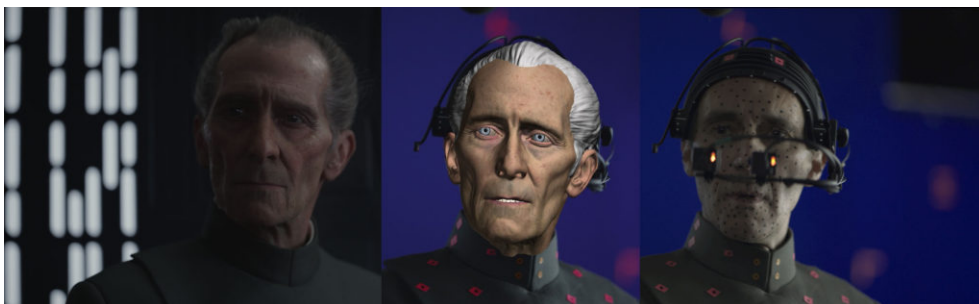
Animate strange creatures ([fxguide.com](https://www.fxguide.com))

Face Pipeline: Snap Solver “Fein”



Correct errors ([fxguide.com](https://www.fxguide.com))

Face Pipeline: Manuelle Animation



Manually correct for differences ([fxguide.com](https://www.fxguide.com))

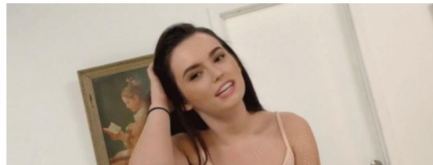
DeepFakes

We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now

A user-friendly application has resulted in an explosion of convincing face-swap porn.

SHARE TWEET

Samantha Cole
Jan 24 2018, 7:13pm



📺 Motherboard: We Are Truly Fucked: Using Tensorflow to superimpose faces.

Video 14.16:



📺 You Won't Believe What Obama Says In This Video! (YouTube, 1:12)

📺 Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news (The Verge)

6 Bildnachweis

Alle Abbildungen, wenn nicht anders angegeben, aus:

Malaka, Rainer; Butz, Andreas; Hussmann, Heinrich: *Medieninformatik – Eine Einführung*. ISBN 978-3-8273-7353-3, München: Pearson Studium, 2009.

Images

Copyright by the respective photographers:

- Slide 70: Paper illustration by M. Weiser
- Slide 75: Laptop Keyboard by DeclanTM, <http://www.everystockphoto.com/photo.php?imageId=3056657>
- Slide 75: Mobile phones by S. Khabbazzadeh
- Slide 75: Bowl of glass by joygant, <http://www.flickr.com/photos/joygant/1291048974/>

References

Literatur

Ken Ducatel, Marc Bogdanowicz, Fabiana Scapolo, Jos Leijten, and Jean-Claude Burgelman. ISTAG scenarios for ambient intelligence in 2010. Technical report, IST Advisory Group, 2001.

Rainer Malaka, Andreas Butz, and Heinrich Hussmann. *Medieninformatik – Eine Einführung*. Pearson Studium, Munich, 2009.

Mark Weiser. The computer for the 21st century. *Scientific American*, pages 94–104, September 1991.

Mark Weiser. Some computer science issues in ubiquitous computing. *Communications of the ACM*, 36(7):75–84, July 1993. ISSN 0001-0782. doi: 10.1145/159544.159617. URL <http://doi.acm.org/10.1145/159544.159617>.