

# Zusammenfassung

Stand: 5. Februar 2020

Jörg Cassens

Medieninformatik

WS 2019/2020



## 1 Einleitung

### Prüfungstermine

- Zwei gleichwertige Prüfungstermine
- Zu Beginn und am Ende der vorlesungsfreien Zeit
  - Dienstag, 11.02.2018.
    - \* 10:00 Uhr
    - \* A 9, Samelsonplatz
  - Dienstag, 31.03.2018.
    - \* 10:00 Uhr
    - \* A 9, Samelsonplatz
- Länge je nach Credits
  - 8 ECTS = 120 Minuten (IMIT, WI, ...)
  - 6 ECTS = 90 Minuten (LA, ...)
- Verbindliche Anmeldung im POS/LSF erforderlich (eine Woche vor der Prüfung)
- Absprachen zum Nachteilsausgleich bitte bis eine Woche vor dem jeweiligen Prüfungstermin
- Beantwortung auf Deutsch oder Englisch

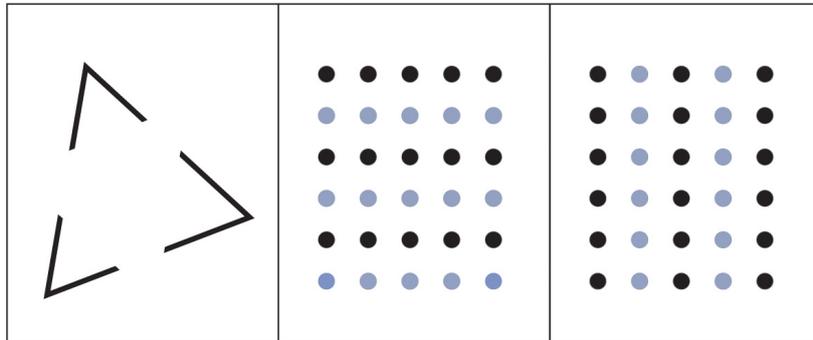
### Thema

- Überblicksvortrag
  - Nicht erschöpfend
  - Anleitend
- Fragen...
  - ... wurden teilweise in Übungen behandelt
  - ... können im learnweb diskutiert werden
- Danke...
  - ... vorab für gegenseitige Beantwortung von Fragen
  - ... für Klausurfragen

## 2 Überblick

### Visuelle Wahrnehmung

- Viele Dinge sehen wir eigentlich mit dem Gehirn



**Abbildung 1.2:** Gestaltgesetze. Links: Offene Formen werden zu einem geschlossenen Dreieck ergänzt. Mitte und rechts: Je nachdem, welche Punkte schwarz und grau sind, neigen wir dazu, eher Zeilen oder Spalten zu erkennen.

### Gestaltgesetze (-prinzipien)

- Gesetz der **Nähe**
  - gruppiert Dinge zusammen, die räumlich oder zeitlich nah sind
- Gesetz der **Ähnlichkeit/Gleichheit**
  - gruppiert Bildteile, die nach Farbe, Form, Helligkeit, Größe, Orientierung ähnlich sind
- Gesetz der **guten Fortsetzung**
  - präferiert räumliche oder zeitliche Einfachheit
- Gesetz der **Geschlossenheit**
  - neigt dazu, kleine Lücken aufzufüllen
- Gesetz des **gemeinsamen Schicksals**
  - Objekte werden gerne als Gruppen wahrgenommen
- Weiterhin komplexere Prinzipien
  - Gesetz der **Symmetrie**
  - Unterscheidung von **Vorder- und Hintergrund**

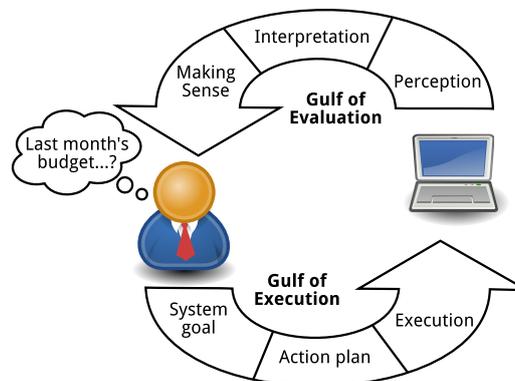
### Das Auditive System

- Beim Hören nehmen wir mechanische Wellen wahr
  - Außenohr: schützt, verstärkt Klang
  - Mittelohr: leitet Vibrationen ans Innenohr
  - Innenohr: frequenzspezifische Sinnesnerven übertragen Reize
- Wahrnehmung etwa 18 bis 20.000 Hz (meist weniger Höhen)
- deutliches Nachlassen mit dem Alter
- Räumliche Auflösung deutlich schlechter als beim visuellen System – mehrere Grad
- Zeitliche Auflösung deutlich besser, Reize können schon bei 2 bis 3 ms unterschieden werden
- Wenige verschiedene Sensorzellen

## Beispiele für Medientheoretiker

- **Marshall McLuhan:** "The Medium is the Message"
  - Inhalte sind nicht neutral zu ihren Trägern
  - Materialität und die Information sind gegenseitig abhängig
  - Auch: Medien als "Extensions of Man" zur Unterstützung und Anreicherung menschlicher Fähigkeiten
- **Jean Baudrillard:** "Simulacra and Simulation"
  - Verschwinden des Realen zugunsten von Simulation und Hyperrealität
  - Zeichenwelten, Zwischenwelten, Simulationen treten an Stelle konkreter Realität
- **Neil Postman:** "Amusing ourselves to Death"
  - Sozialkritische Auseinandersetzungen mit der Bedeutung und Wirkung von vor allem elektronischen Medien
  - Kritik des (amerikanischen) Bildungssystems und des computerunterstützten Lernens
  - Rückbesinnung auf die Narration, Spiritualität und amerikanische Grundwerte

## Computer als Semiotische Maschine



☞ Donald Norman: *The Design of Everyday Things* (1988)

## 3 Geschichte

### Künstliche Medien

- Künstliche Medien wurden geschaffen, um den Menschen zunächst hinsichtlich seines Körpers und später auch im Geist zu unterstützen.
  - Faustkeil
  - Rechenhilfen
- McLuhan: "Extensionen des Menschen"
  - Sollen dem Menschen helfen, auf die Welt effektiver und effizienter einzuwirken oder einfacher am sozialen und kulturellen Leben teilnehmen zu können
- Entwicklung verläuft nicht gleichmäßig
  - **Evolutionäre Phasen** verbessern vorhandene Lösungen durch kleine Entwicklungsschritte, die meist durch sorgfältige Analysen vorbereitet und anschließend umgesetzt werden
  - **Revolutionäre Phasen** stellen die existierenden Lösungen grundlegend in Frage und ersetzen oder ergänzen diese durch neue Lösungen

## Vannevar Bush: MEMEX

- MEMEX (1945): Vannevar Bush
- “The human mind [...] operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain.” (V. Bush, 1945, S.101)
- Merkmale:
  - Basis: Mikrofilme mit schnellem Archivzugriff
  - Möglichkeit zur Textverknüpfung
  - Erstellung von eigenen Einträgen
  - elektronische Kameras am Kopf
  - Sprachaufnahme mit Zeitstempel zur Synchronisation
  - arithmetische und logische Manipulation von Daten
  - Browser-Prinzip (History, Bookmarks) zur gemeinsamen Nutzung

## MEMEX: Vision

### Vision

“A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.

It consists of a desk, and while it can presumably be operated from a distance, it is primarily the piece of furniture at which he works. On the top are slanting translucent screens, on which material can be projected for convenient reading. There is a keyboard, and sets of buttons and levers. Otherwise it looks like an ordinary desk.”

- nichtsequentielles, assoziatives Katalogisieren (und Finden) von Daten (später nach Nelson: Hypertext)
- “Selection by association, rather than indexing”

## Dartmouth Workshop (1956)

### Proposal

“... that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.”

A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence



(Proposal: McCarthy et al, Photos: Margaret Minsky, James Moor)

## Ivan Sutherland: Sketchpad

- andere schreiben und reden, Sutherland programmiert
- andere interpretieren seine Arbeit
- Sketchpad (1963)
  - z.B. master drawing und instances (Vorläufer zu Objektorientierung)
  - interaktive 3-dim. Manipulation durch getrennte Koordinatensysteme
  - Dissertation am MIT
  - computer-aided Design (CAD)

- Basis vieler Graphik-Konstruktions- und Editier-Funktionen

-  Videovortrag Alan Kay [04:05 - 09:20 Min.]



Bildquellen:  Digital Humanities Quarterly,  computerhistory.org

### Ivan Sutherland: The Ultimate Display

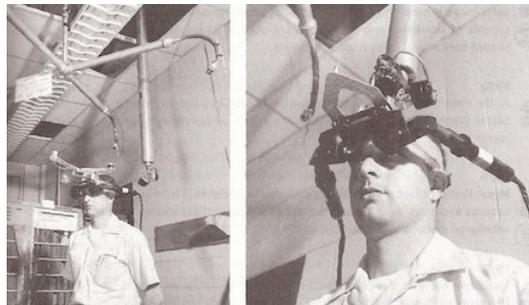
“The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal. With appropriate programming such a display could literally be the Wonderland into which Alice walked.”

(Ivan Sutherland (1965), The Ultimate Display)

- A virtual world viewed through a HMD and appeared realistic through augmented 3D sound and tactile feedback.
- Computer hardware to create the virtual world and maintain it in real time.
- The ability users to interact with objects in the virtual world in a realistic way

### Ivan Sutherland: Sword of Damocles

- Sword of Damocles (1968)
- Erstes AR/VR Head Mounted Display, die mit einem Computer statt einer Kamera verbunden war
- Primitive Wireframe-Objekte und Räume



A head mounted three dimensional display, Ivan Sutherland, 1968

### Douglas Engelbart: NLS

- NLS (oNLine System):
  - Software für Time-Sharing Computer um intellektuelle Zusammenarbeit zu unterstützen
  - Bildschirme, Keyboard, Maus!
  - Hierarchisches Informationsmanagement für Text, Diagramme, Bilder, Video-Konferenz
  - Hypertext
- 1968 Fall Joint Computer Conference in San Francisco:
  - Ergebnis von 4 Jahren Arbeit in 90 Minuten
  - Standing Ovationen von 3000 Zuschauern
-  Videovortrag Alan Kay [09:20 - 19:15 Min.]



## Bolt: Put That There (1980)

- Richard Bolt (MIT)
- Multimodale Interaktion
  - Gestensteuerung
  - Spracheingabe und -ausgabe
  - Video-Ausgabe



Put That There (5:58)

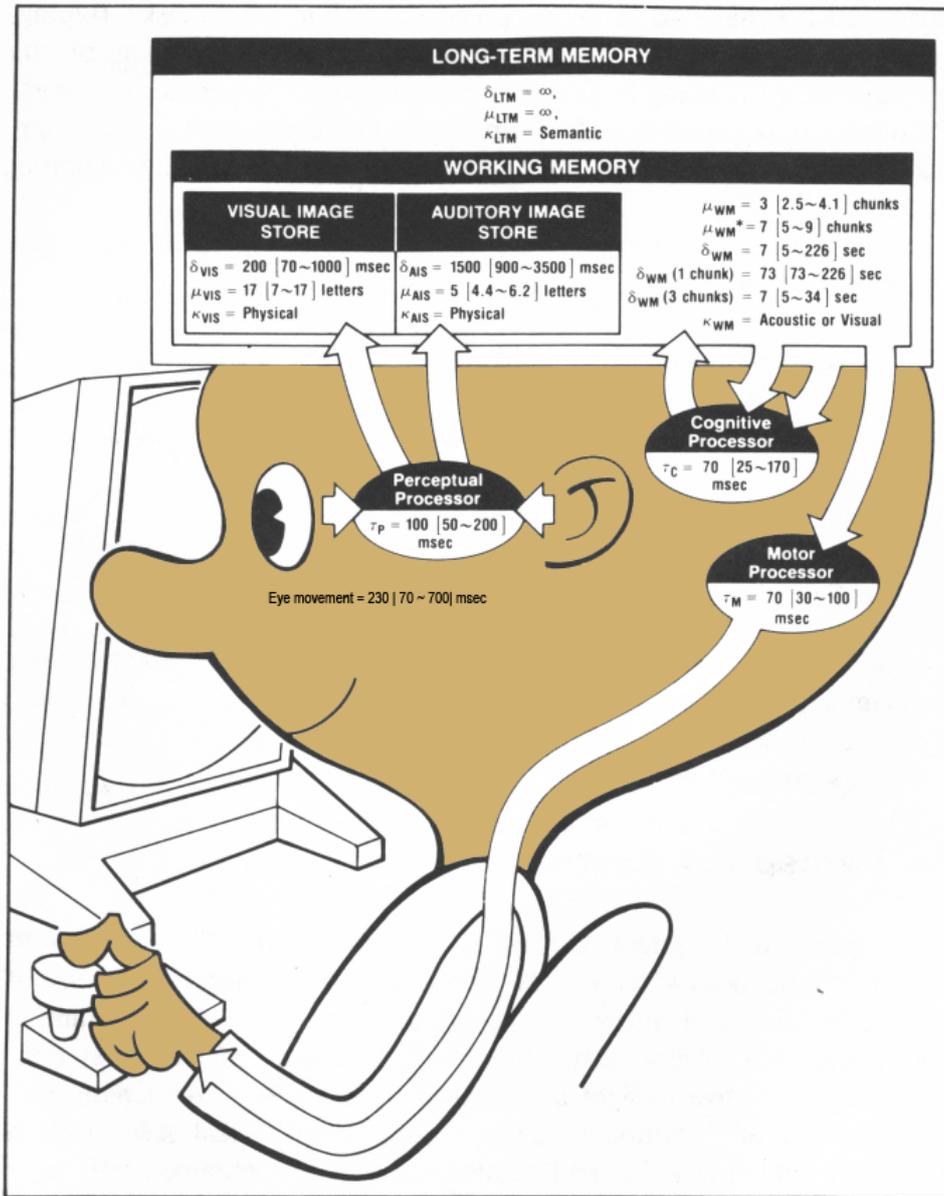
## Tangible Computing (2005)

- Tangible Media
- I/O Brush is a new drawing tool to explore colors, textures, and movements found in everyday materials by “picking up” and drawing with them.
- Ryokai, Marti & Ishii, MIT Media Lab 2005
- [Wayback Machine](#), [Video](#)



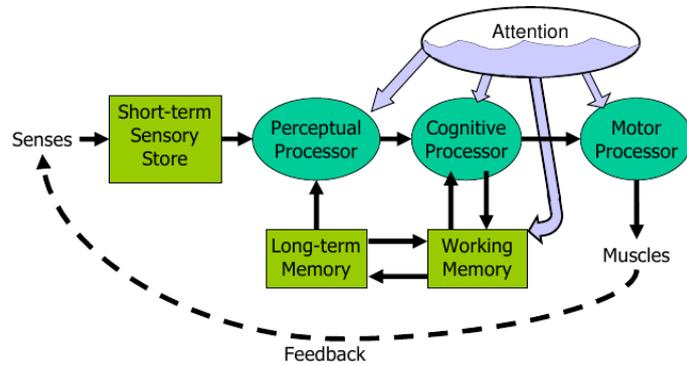
## 4 Human Capabilities

### Model Human Processor (MHP)



Card, Newell & Moran (1983)

**Human Information Processing (HIP)**



Robert Miller (2004)

**Memory**



- **Encoding:** type of things stored
- **Size:** number of things stored
- **Decay time:** how long memory lasts

### Short-Term Sensory Store

- Visual information store
  - encoded as physical image (curves, edges, length – not as pixels)
  - size ~ 17 [7-17] letters (convenient signals, not signs)
  - decay ~ 200 ms [70-1000 ms]
- Auditory information store
  - encoded as physical sound
  - size ~ 5 [4.4-6.2] letters
  - decay ~ 1500 ms [900-3500 ms]
- Both are preattentive: they don't need the spotlight of attention to focus on them in order to be collected and stored
- Attention can be focused on the visual or auditory stimulus after the fact: "What did you say? Oh yeah."

### Working Memory (WM)

- Working memory is where you do your conscious thinking
- Working memory is where the cognitive processor gets its operands and drops its results
- Small capacity:  $(4 \pm 2) - (7 \pm 2)$  "chunks"
  - This number is often quoted
  - Empirical evidence can be interpreted in different ways
- Fast decay (7 [5-226] sec)
- Maintenance rehearsal fends off decay
- Interference causes faster decay

### Long-term Memory (LTM)

- Probably the least understood part of human cognition
- It contains the mass of our memories
- Huge capacity
- Little decay
- Apparently not intentionally erased; they just become inaccessible
- Maintenance rehearsal (repetition) appears to be useless for moving information into long-term memory
- Elaborative rehearsal moves chunks from WM to LTM by making connections with other chunks
- Compare e.g. mnemonic techniques like associating things you need to remember with familiar places, like rooms in your childhood home

## Processors

- Processors have a cycle time
  - $T_p \sim 100\text{ms}$  [50-200 ms]
  - $T_c \sim 70\text{ms}$  [30-100 ms]
  - $T_m \sim 70\text{ms}$  [25-170 ms]



- Fastman may be 10x faster than Slowman; Middleman is typical (named by Card, Newell, Moran)
- Variations not only between individuals, but also depending on conditions: slow reading in the dark, fast processing when playing WoW

## Bottom-up vs. Top-Down Perception

- Bottom-up uses features of stimulus
  - Identifying features
- Top-down uses context of perception
  - temporal in auditory perception
  - spatial in visual perception
  - draws on long-term memory

TAE CAT

- H and A are represented by the same shape, but can be distinguished because of their context

## Chunking

- “Chunk”: the unit of perception or memory
- Chunking depends on presentation and what you already know
  - defined symbols or activated past experience

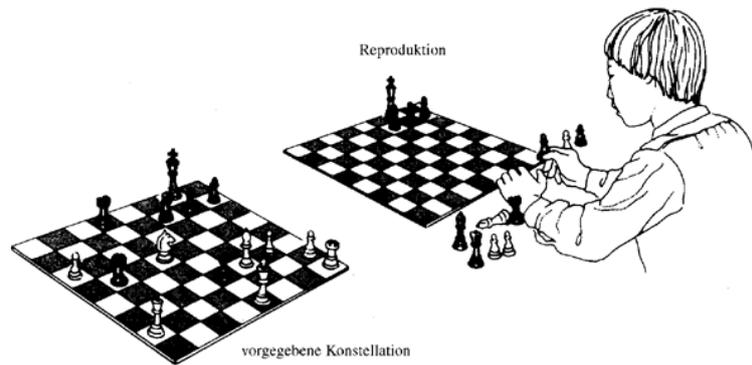
B M W S A P A O L I B M F B I  
(15 chunks)

MWS APA OLI BMF BIB  
(still 15 chunks to most people)

BMW SAP AOL IBM FBI  
(5 chunks to most)

## Chess: Experts vs. Novices

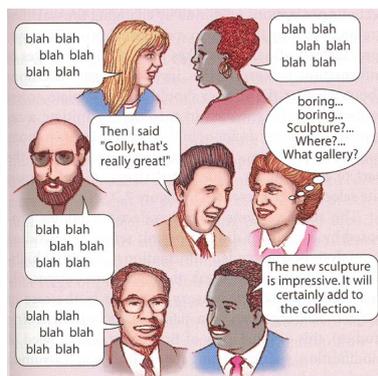
Chess masters are better than novices at remembering real game configurations, same performance on random boards



Reproduction task by Chase und Simon (1973)(in Anderson 2001, S.301).

## Übung: Interferenz

- Spotlight metaphor: You can focus your attention (and your perceptual processor) on only one input channel in your environment at a time.
- Beispiel: Cocktailparty-Effekt (Cherry, 1953)



## Dichotic Listening Task



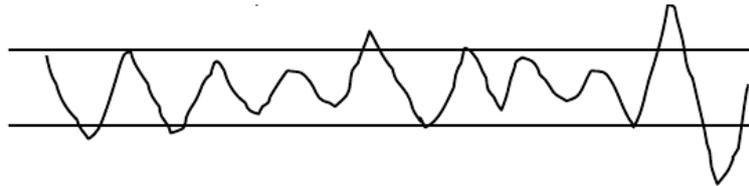
## Cognitive Processing

- Cognitive processor
  - compares stimuli
  - selects a response
- Types of decision making (Rasmussen)
  - Skill-Based Behaviour
    - \* Automatic reaction to sensory input
    - \* Breaking lights – breaking
  - Rule-Based Behaviour
    - \* Based on sensory input, rules are fired
    - \* Happens when there is no automatic response
    - \* Choice of rule based on signs recognized
    - \* Regulating speed and direction when exiting a freeway
  - Knowledge-Based Behaviour
    - \* Conscious problem solving
    - \* Happens when there are no rules
    - \* Triggered by interpreted symbols
    - \* Stuttering motor – continue or stop?

## Motor Processing I

- Open-loop control
  - Motor processor runs a program by itself
  - cycle time is  $T \sim 70$  ms
- Closed-loop control
  - Muscle movements (or their effect on the world) are perceived and compared with desired result
  - cycle time is  $T_c + T_p + T_m \sim 240$  ms

## Motor Processing II



- The frequency of the sawtooth carrier wave is dictated by open-loop control
- The frequency of the wave's envelope, the corrections to be made to get the scribble back to the lines, is closed-loop control

# 5 Kanäle, Codecs und Medien

## Ebenen

Wir können Medien auf verschiedenen Ebenen unterscheiden:

- **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)?

## Ebenen: Beispiele

- **Präsentation & Aufnahme – Medial**
  - Radio: monomedial
  - TV: multimedial
  - Stummfilmkamera: 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

## Analog und Digital

**Analoges Signal:** Deterministische und kontinuierliche Änderung einer physikalischen Größe entsprechend eines Meßwertes der zu übertragenden Größe.

**Digitales Signal:** Annäherung an den Meßwert in einem festen Raster über die Zeit oder den Raum mittels eines endlichen Vorrats möglicher Repräsentationen.

## Abtastung

- Jedes Signal, welches am Ende der Verarbeitungskette als mediales Angebot präsentiert wird, muß irgendwann in digitale Form gebracht werden
  - Synthese: das Ausgangsmaterial wird in digitaler Form hergestellt
  - Abtastung: ein analoges Signal wird in ein digitales umgewandelt
- Abtastung:
  - Wie *häufig*? → Sampling, Diskretisierung
  - Wie *genau*? → Quantisierung

## Diskretisierung und und Quantisierung

**Diskretisierung:** Festes Raster von Meßpunkten gleichen Abstands auf der Achse oder den Achsen, auf der sich das analoge Signal verändert. Die Dichte wird als Abtastrate bezeichnet. Zu jedem Meßpunkt wird jeweils das Signal (Sample) bestimmt.

**Quantisierung:** Darstellung der ermittelten Meßwerte in einem festen Werteraster. Die Zahl der zur Verfügung stehenden Zahl der Bits wird auch als Auflösung der Digitalisierung bezeichnet.

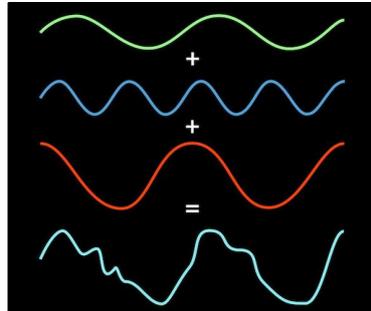
## Abtasttheorem

1. Ein reales Signal kann als Überlagerung verschiedener Grundsignale verschiedener Frequenzen aufgefaßt werden.
2. Für die Wahl der Abtastrate ist die Frequenz des im digitalen Signal enthaltenen Anteils mit der höchsten Frequenz entscheidend.

**Abtasttheorem** (Shannon, Nyquist, Whittaker, Kotelnikow): Wenn ein kontinuierliches Signal mit einer oberen Grenzfrequenz  $f_{max}$  mit einer Abtastrate von mehr als  $2 * f_{max}$  abgetastet wird, kann man das Ursprungssignal ohne Informationsverlust aus dem abgetasteten Signal rekonstruieren.

## Fourier-Transformation

- Jean Baptiste Joseph Fourier (1768-1830)
- Jede periodische Schwingung kann durch eine (unendliche) Summe von überlagerten Cosinus-Schwingungen angenähert werden
- Überlagerung harmonischer Schwingungen



Christine Daniloff/MIT

## Aliasing

- Bei der Rekonstruktion können Signalanteile entstehen, die im Originalsignal nicht vorhanden waren
- Entsteht z.B. durch eine zu kleine Abtastrate
- Eine theoretisch mögliche vollständige Rekonstruktion kann durch Eigenschaften des Ausgabegerätes zu Aliasing-Effekten führen (Moiré-Effekt bei Graphik auf kleiner Auflösung)
- Abhilfe: Ausfiltern zu hoher Frequenzanteile ("Weichzeichner")



## Information und Repräsentation

- Information ist ein nichtstoffliches Phänomen, das durch die Interpretation bestimmter Zeichen oder Zeichenfolgen entsteht, die man die Repräsentation der betreffenden Information nennt.
- Informationstheorie nach Shannon analysiert Codierungen vor allem unter dem Aspekt des Auftretens eines bestimmten Zeichen eines Zeichenvorrats (stochastischer Ansatz)
- Grundlegend: Nachricht und Nachrichtenquelle.
- In der Informatik ist besonders die Repräsentation im Binärformat relevant, also als Abfolge von 0 und 1

## Ausgangspunkt

Wir haben eine Codierung von Informationen in einem beliebigen Ausgangsformat und suchen eine möglichst effiziente Codierung im Binärformat

## Definitionen

**Zeichenvorrat:** Endliche Menge von Zeichen.

**Nachricht:** Eine Nachricht im Zeichenvorrat  $A$  ist eine endliche Sequenz von Zeichen  $a$  aus  $A$ .

**Codierung:** Seien  $A$  und  $B$  Zeichenvorräte. Dann ist eine Codierung  $c$  von  $A$  in  $B$  eine Abbildung von Nachrichten aus  $A$  in Nachrichten aus  $B$ :  $c(a) \rightarrow b, a \in A, b \in B$ .

Im folgenden betrachten wir nur die Einzelzeichencodierung.

**Nachrichtenquelle:** Eine Nachrichtenquelle nach Shannon ist ein Zeichenvorrat  $A$  zusammen mit einer Wahrscheinlichkeitsverteilung, die für jedes Zeichen  $a \in A$  die Wahrscheinlichkeit seines Auftretens angibt.

Wahrscheinlichkeit von  $a \in A$  wird mit  $p_a$  bezeichnet,  $0 \leq p_a \leq 1, \sum p_a = 1$

## Beispiel Nachrichtenquelle I

Zeichen $a$	A	B	C	D
Wahrscheinlichkeit $p_a$ in Quelle 1	1.0	0.0	0.0	0.0
Wahrscheinlichkeit $p_a$ in Quelle 2	0.25	0.25	0.25	0.25
Wahrscheinlichkeit $p_a$ in Quelle 3	0.5	0.25	0.125	0.125

Tabelle 2.1: Drei Beispiele für Nachrichtenquellen

- Unterschiedlicher Informationsgehalt der einzelnen Zeichen
  - Quelle 1: Nur A, bringt keine weiteren Informationen
  - Quelle 2: Jedes Zeichen bringt gleichen Informationsgehalt entsprechend der Auswahl eines Elements aus einer vierelementigen Menge
  - Quelle 3: A bringt weniger Informationen, es ist "weniger überraschend"
- Informationsgehalt beruht ausschließlich auf der Wahrscheinlichkeit des Auftretens, keine andere Verständnisebene notwendig

## Beispiel Nachrichtenquelle II

- Wie kann man den Informationsgehalt messen?
- 1 Bit ist die Informationsmenge, die notwendig ist, um aus zwei Elementen eines auszuwählen (Entscheidungsgehalt)
  - 2 Bit = vier Elemente, 3 Bit = acht Elemente, ...

Zeichen $a$	A	B	C	D
Entscheidungsgehalt in Quelle 1 [Bit]	0	undefiniert	undefiniert	undefiniert
Entscheidungsgehalt in Quelle 2 [Bit]	2	2	2	2
Entscheidungsgehalt in Quelle 3 [Bit]	1	2	3	3

Tabelle 2.2: Beispiele für den Informationsgehalt von Zeichen

## Wortlänge

- Wir sind in erster Linie an binären Codierungen interessiert
- Wenn wir wissen wollen, wieviel Platz die Codierung eines einzelnen Zeichens im Durchschnitt belegt, können alle möglichen Codierungen unter Einbeziehung der Auftrittswahrscheinlichkeit eines Zeichens betrachten

Die **durchschnittliche Wortlänge**  $L$  ist die nach Auftrittswahrscheinlichkeiten gewichtete Summe der Wortlängen aller Codierungen:

$$L = \sum_{a \in A} p_a * |c(a)|$$

## Beispiel

Zeichen $a$	A	B	C	D
Wahrscheinlichkeit $p_a$ in Quelle 3	0.5	0.25	0.125	0.125
Codierung $c_1$	<b>00</b>	<b>01</b>	<b>10</b>	<b>11</b>
Wortlänge	2	2	2	2
Durchschnittliche Wortlänge $L = 0.5 \cdot 2 + 0.25 \cdot 2 + 0.125 \cdot 2 + 0.125 \cdot 2 = 2$				

Tabelle 2.4: Beispiel für eine redundante Codierung

Zeichen $a$	A	B	C	D
Wahrscheinlichkeit $p_a$ in Quelle 3	0.5	0.25	0.125	0.125
Codierung $c_2$	<b>0</b>	<b>10</b>	<b>110</b>	<b>111</b>
Wortlänge	1	2	3	3
Durchschnittliche Wortlänge $L = 0.5 \cdot 1 + 0.25 \cdot 2 + 0.125 \cdot 3 + 0.125 \cdot 3 = 1.75$				

Tabelle 2.5: Beispiel für eine optimale Codierung

## Entropie

- Entropie ist das "Maß der Unordnung"
- Quelle 2 ist chaotisch, alle Zeichen treten gleich häufig auf
- Quelle 3 gibt hingegen das A bevorzugt aus

Die **Entropie** einer Nachrichtenquelle ist der durchschnittliche Entscheidungsgehalt eines Zeichens der Nachrichtenquelle:

$$H = \sum_{a \in A} p_a * x_a = \sum_{a \in A} p_a * \log_2 \left( \frac{1}{p_a} \right)$$

## Redundanz und Optimale Codierung

**Redundanz:** Die Redundanz  $R$  einer binären Codierung für eine Informationsquelle ist die Differenz der mittleren Wortlänge und der Entropie:

$$R = L - H$$

Eine **Codierung** einer Nachrichtenquelle heißt **optimal**, wenn die Redundanz der Codierung gleich Null ist.

## Huffman-Codierung

- Universelle, verlustfreie Kompression
- Vertreter einer Gruppe von Verfahren, die auf Shannon's informationstheoretischen Überlegungen basiert
- Statistische oder stochastische Verfahren
- Ausgangspunkt: Shannonsche Informationsquelle mit einer Angabe von Auftrittswahrscheinlichkeiten
- Der Huffman Algorithmus konstruiert zu einer Nachrichtenquelle eine Codierung, die unter bestimmten Bedingungen optimal ist

Erster Durchlauf				
Zeichen	A	B	C	D
Wahrscheinlichkeit	0,5	0,25	0,125	0,125
Zweiter Durchlauf				
Zeichen/Baum	A	B		
Wahrscheinlichkeit	0,5	0,25	0,25	
Dritter Durchlauf				
Zeichen/Baum	A			
Wahrscheinlichkeit	0,5	0,5		
Vierter Durchlauf				
Zeichen/Baum				
Wahrscheinlichkeit	1,0			

Tabelle 2.7: Ablauf des Huffman-Algorithmus

## Fano-Bedingung

- Problem: variable Codelänge
- Lösungsmöglichkeit
  - Trennzeichen
  - Aufbau (Morse Code: Pausen)
- Hier: Fano-Bedingung

**Fano-Bedingung:** Eine Codierung  $C$  eines Zeichenvorrats  $A$  in einen Zeichenvorrat  $B$  erfüllt die Fano-Bedingung, wenn für alle Zeichen  $x$  und  $y$  aus  $A$  gilt, daß das Wort  $c(x)$  nicht Anfang des Wortes  $c(y)$  ist, also keiner der verwendeten Codes Anfang eines anderen ist.

## Arithmetische Codierung

- Huffman: Ideal wenn Wahrscheinlichkeiten Kehrwerte von Zweierpotenzen
- Entwicklung von Verfahren, die auf beliebigen Wahrscheinlichkeitsverteilungen gute Ergebnisse liefern
- Grundidee Arithmetische Codierung: Ordne einer Nachricht ein Teilintervall aus den reellen Zahlen zwischen 0 und 1 zu
- Länge des Intervalls eine Funktion der Wahrscheinlichkeit des Auftauchens
- bei fortschreitender codierung werden fortlaufend kleiner Intervalle gebildet
- Wahrscheinlichkeiten der Nachrichtenquelle ergeben erste Intervalleinteilung

## Laufmängencodierung

- Beispiel für zeichenorientierte, universelle verlustfreie Kompression
- Unabhängig von informationstheoretischen Überlegungen
- Grundidee: Speicherung der Wiederholungen eines Zeichens
- AABBBBBBEEDDDDDDDDDDB  $\rightarrow$   $\langle A,2 \rangle \langle B,5 \rangle \langle E,2 \rangle \langle D,11 \rangle \langle B,1 \rangle$
- Verschiedene Möglichkeiten der Speicherung
  - Dedizierte Trennzeichen
  - Ausnutzen Byte-Struktur (1. Byte Zeichen, 2. Byte Anzahl)

## LZW

- Lempel-Ziv-Welch-Codierung
- Adaptives Verfahren
  - Starte mit initialem Model
  - Lese die Nachricht
  - Codiere die Nachricht und aktualisiere das Model
- Hier: Wörterbuch für in einer Nachricht vorkommende Teilworte
  - Wörterbuch ist durchnummeriert
  - Nummer übertragen
- Algorithmus speichert nur Teilzeichenfolgen im Wörterbuch, die notwendig sind
- gleichzeitig passiert die Spaltung der Nachricht

### LZW: Ablauf

1. Initialisiere das Wörterbuch mit allen Zeichenketten der Länge 1 und starte mit dem Eingabetext an Position 1
  2. Finde im Wörterbuch den längsten String W der ab der Startposition in der Eingabe auftaucht
  3. Gebe den Wörterbuch-Index für W aus und entferne W aus der Eingabe
  4. Füge W plus dem nächsten Zeichen in der Eingabe dem Wörterbuch hinzu
  5. Gehe zu Schritt 2
- In der Regel maximale Länge des Wörterbuchs festgelegt
  - Codes länger als zu codierende Zeichen (Länge in Bits), Kompression durch Wiederholung von Substrings

### LZW: Dekompression

- Nur das Ausgangs-Wörterbuch muß bekannt sein
    - Texte: ASCII?
  - Der Rest des Wörterbuches wird bei der Dekompression mit erstellt
1. Lese das nächste Zeichen
  2. Sehe im Wörterbuch nach und gebe das Zeichen aus
  3. Erster Buchstabe des gelesenen Teilwortes wird mit gemerktem Teilwort zu einem neuen Eintrag
  4. Dekodiertes Teilwort wird gemerkt
  5. Gehe zu Schritt 1

### LZW: Dekompression II

- Die Situation, daß ein Eintrag noch nicht im Wörterbuch vorhanden ist, kann nur in einem Fall auftreten: Eingabestring der Form abababa (Wiederholung)
- Das zuletzt eingetragene Teilwort (aba) wurde gleich nach Erstellung des Eintrags im Wörterbuch benutzt
- Wir nehmen unseren gemerkten Teilstring (ab) und fügen das erste Zeichen an das Ende an (aba), geben den Teilstring aus und fügen ihn im Wörterbuch hinzu

**Wörterbuch:** 0:a, 1:b, 2:ab, 3:ba, 4:aba,

Eingabe	?	mem	WB	Ausgabe
0	0=a			a
0,1	1=b	a	2:ab	ab
01,2	2=ab	b	3:ba	abab
012,4	4=aba	ab	4:aba	abababa

## LZW: Beispiel

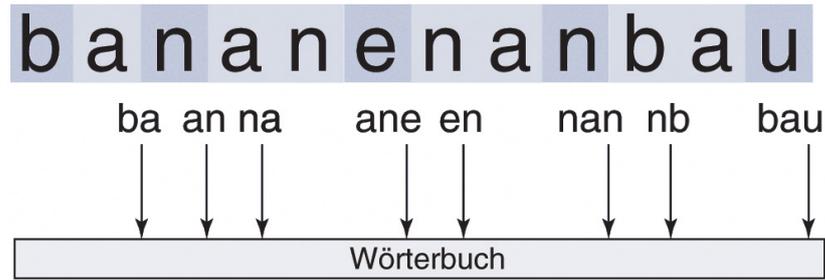


Abbildung 2.9: Ablauf des LZW-Algorithmus an einem Beispiel

### Klassifikation von Kompressionsverfahren I

Eine **Kompressionsverfahren**, das für alle Daten unabhängig von Ursprung und Bedeutung angewendet werden kann heißt **universell**. Verfahren, die nur auf Daten eines bestimmten Typs (Audio, Bild) anwendbar sind heißen **speziell**.

- Universell: zip, gzip, bzip2
- Speziell: z.B. Ausnutzen von Eigenschaften, wie die daß es in Photos häufig Flächen gleichen Farbtons gibt.

### Klassifikation von Kompressionsverfahren II

Eine **Kompressionsverfahren** heißt **verlustbehaftet**, wenn bei der Kompression Informationen aus dem Original verloren gehen. Im Gegensatz stehen **verlustfreie** Methoden, die die vollständige und genaue Rekonstruktion des Originals ermöglichen.

	Universelle Verfahren	Spezielle Verfahren
Verlustfreie Verfahren	Beispiele: Huffman, LZW	Beispiele: PNG, AIFF
Verlustbehaftete Verfahren	(nicht sinnvoll)	Beispiele: JPEG, MP3

Tabelle 2.6: Klassifikation von Kompressionsverfahren

## 6 Audio

### Hörbare Frequenzen

- Der vom Menschen hörbare Frequenzraum
- Beachte: für Sprache reicht ein weitaus kleinerer Teil aus

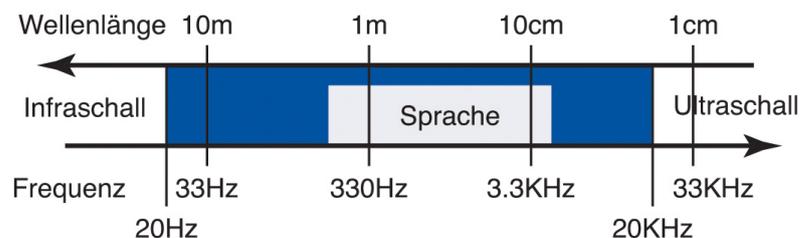
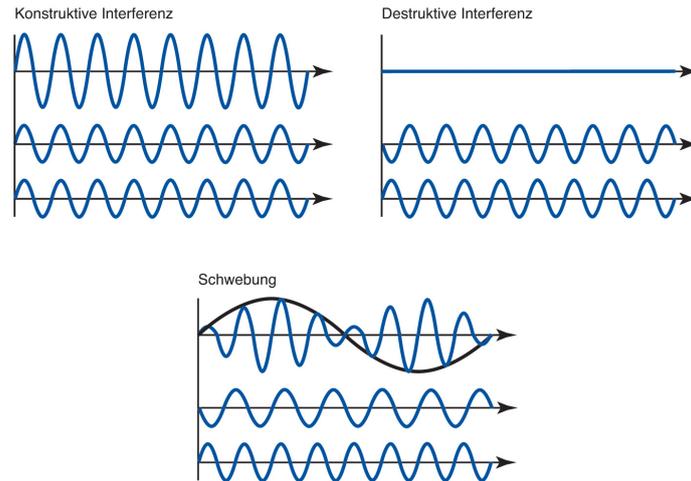


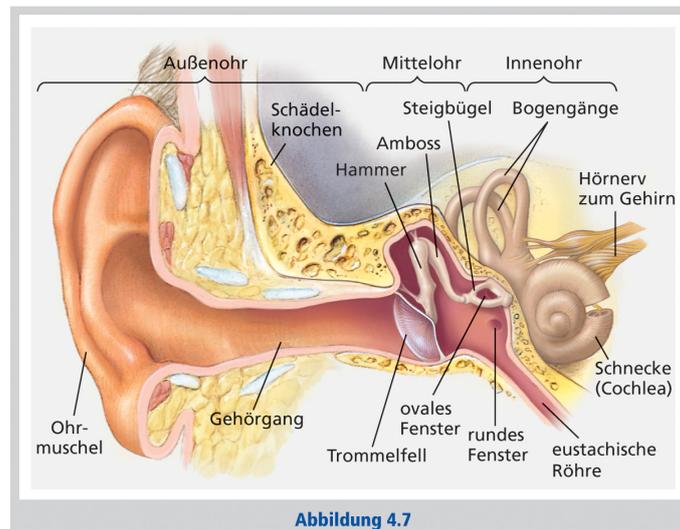
Abbildung 4.2: Hörbarer Frequenzbereich: Sprache nutzt nur einen Teil der hörbaren Frequenzen aus.

## Addition



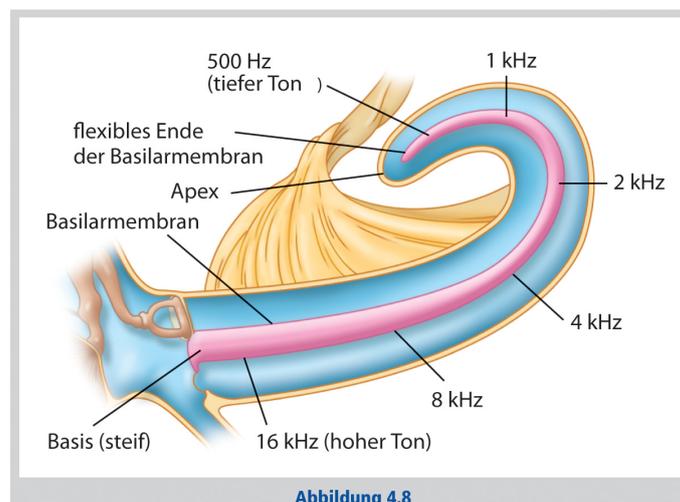
**Abbildung 4.6:** Konstruktive Interferenz, destruktive Interferenz und Schwebung (abwechselnd konstruktive und destruktive Interferenz)

## Akustische Wahrnehmung



**Abbildung 4.7**

## Akustische Wahrnehmung (contd.)



**Abbildung 4.8**

## Maskierung

- Bei nahe beieinander liegenden Reizen kann ein energiereiches Signal eine nahe liegendes, schwächeres Signal maskieren (Frequenzmaskierung)
- Auch schwache Reize vor oder nach starken Reizen (zeitliche Maskierung)

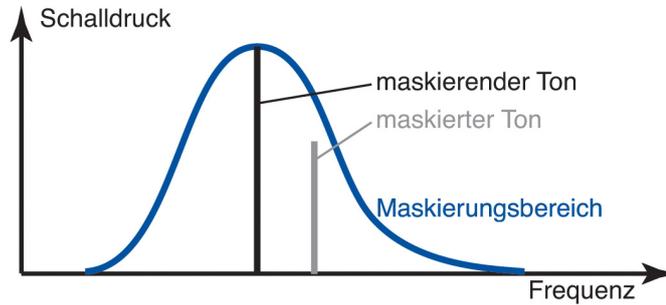


Abbildung 4.9: Maskierung eines leisen Tones durch einen benachbarten lauten Ton

## Lautstärke

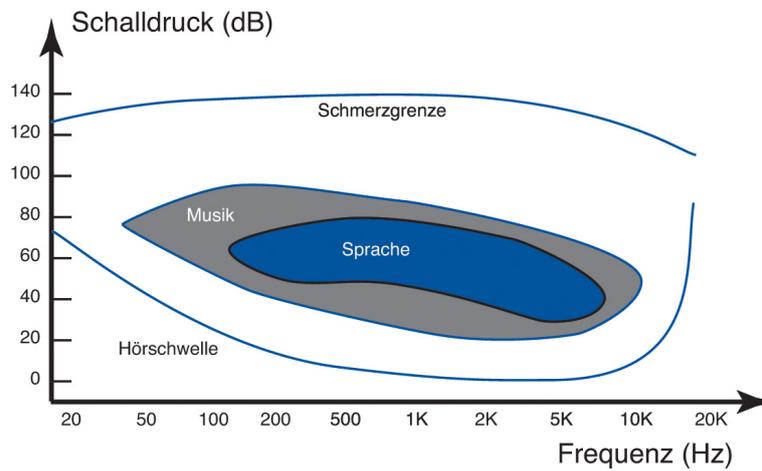
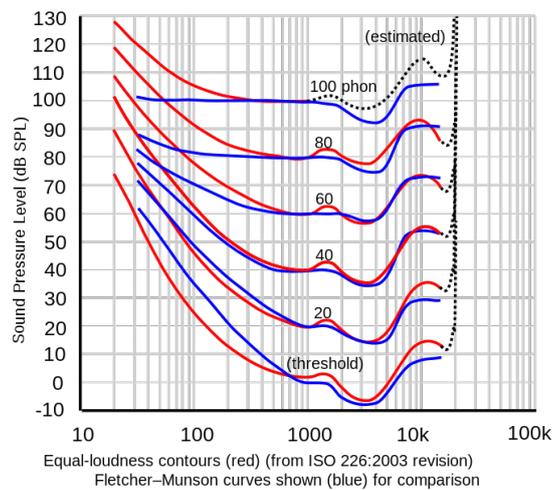


Abbildung 4.10: Lautstärkebereiche des menschlichen Hörens in Abhängigkeit von der Frequenz (vereinfacht)

## Lautstärke und Frequenz



Wikipedia

## Richtungshören

- **Interaurale Zeitdifferenz (ITD):** Unterschied der Laufzeit zwischen beiden Ohren. Gleiche zeitliche Unterschiede führen bei höheren Frequenzen zu höheren Phasenverschiebungen, diese sind besser wahrzunehmen
- **Interaurale Intensitätsdifferenz (IID):** Teilweise Abschwächung der Lautstärke, vor allem bei hohen Frequenzen
- **Head-Related Transfer Function (HRTF):** bei dem Weg um den Kopf werden bestimmte Frequenzanteile verstärkt, andere verschluckt

## Psychoakustik

- Verlustfreie Verfahren führen zu relativ wenig Kompression
- MP3 basiert auf einem psychoakustischen Modell der Tonwahrnehmung
- Nutzt Maskierungseffekte aus
- Empirisch ermittelt
- 27 so genannte kritische Bänder
  - Ein kritische Band ist der Bereich, in dem Töne stärker als ein bestimmter Grenzwert miteinander interagieren
  - Eine solche Interaktion ist z.B. die Maskierung

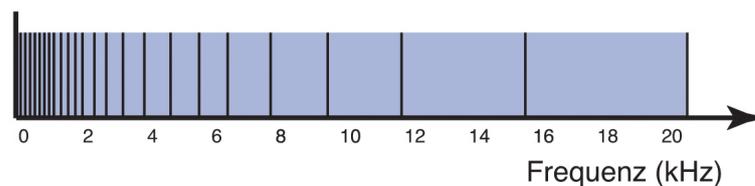


Abbildung 4.12: Kritische Bänder des psychoakustischen Modells

## Kritische Bänder

- Aussagen, welche Frequenzen andere Frequenzen maskieren

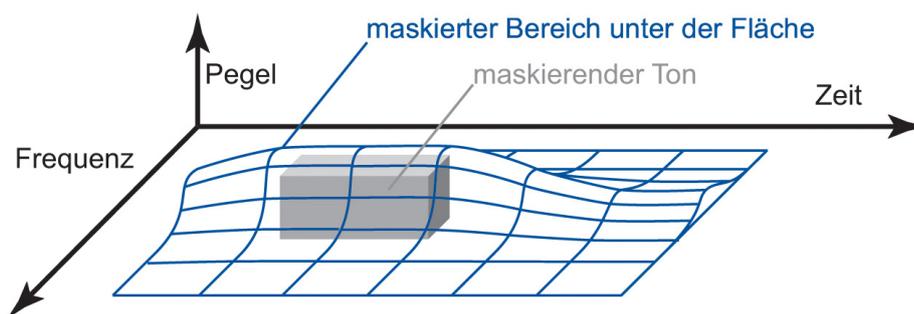
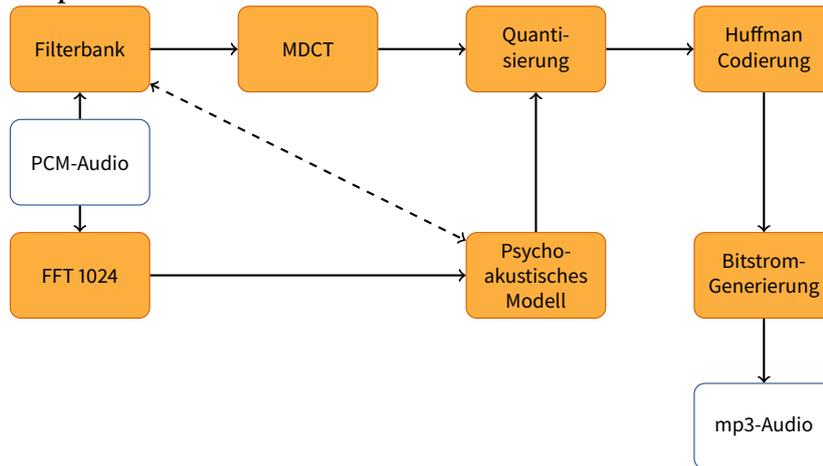


Abbildung 4.13: Frequenzbereich, der durch einen einzelnen Ton räumlich und zeitlich maskiert wird

- Zentrale Idee: Maskierte und Unhörbare Anteile nicht übertragen
- Arbeit im Frequenzraum

## Prinzipschaltbild MP3



Vereinfachtes Prinzipschaltbild nach [Watkinson \[2004\]](#)

## Sampling

- Sehr kurze Zeitabschnitte (36 samples = 0,8 ms)
- Mit Hilfe einer Filterbank in 32 Bänder zerlegt
- “Windowing”: Auswahl von Zeitabschnitten (Frames) für die weitere Codierung
- Unterschiedliche Arten von “Windows” für unterschiedliche Signale
  - “Long Window” bei wenigen Änderungen
  - “Short Window” bei größeren Änderungen
- Danach Modifizierte Diskrete Kosinustransformation (MDCT)
- Darstellung bzgl. 18 Grundfrequenzen innerhalb der jeweiligen Subbänder
- erreichte Darstellung: 32 (Bänder) \* 18 Frequenzen = 576 Basisfrequenzen
- Hohe zeitliche Auflösung

## Frequenzauflösung

- Parallel wird das Eingangssignal mittels schneller Fourier-Transformation in eine Frequenzraumdarstellung bzgl. 1024 Bändern überführt
- Wesentlich niedrigere zeitliche Auflösung, aber bessere Information über Frequenz und Phase
- Dies ist günstiger für die Berechnung der Maskierung
- Auf Basis dieser Zerlegung wird die Maskierung zwischen einzelnen Frequenzen bestimmt
- Aus den Koeffizienten der FFT und dem psychoakustischen Modell wird berechnet, welche Frequenzanteile des zerlegten Signals maskiert und somit nicht wahrgenommen wird

## Verlustbehaftete Codierung

- Genauer: aus den Koeffizienten der FFT und dem psychoakustischen Modell wird berechnet,
  - wie die Quantisierung der Koeffizienten aussieht und
  - welche Window-Größen angewandt werden sollen
- Davon abhängig werden die MDCT-Koeffizienten quantifiziert, wobei viele Nullwerte entstehen
- Die Quantisierung ist auch davon abhängig, welche Bitrate zur Verfügung steht
- Der entstehende Teil wird mit Huffman kodiert
  - Anpassung der Quantisierung an die zur Verfügung stehenden Codetabellen
- Zusätzlich Ausnutzen der Redundanz zwischen Stereokanälen

## FLAC: Vorgehen

1. **Zerlegung:** Bilden von Blöcken von 1000-6000 Samples
2. **Dekorrelierung:** Ausnutzen von Redundanz zwischen Kanälen (rechts/links bei Stereo, weitere Kanäle)
3. **Modellierung:** Predictor für das Audiosignal wird erstellt
  - Verbatim (Kopie)
  - Linear
  - Polynomfunktion oder
  - Linear Predictive Coding
4. **Fehlersignal:** Der Unterschied des Predictors zum tatsächlichen Signal (Residual Signal) wird codiert
  - Entropiebasiert, Rice-Kodierung
5. **Framing:** Alle Kanäle eines Blocks werden mit Metadaten (Fehlererkennung, Timestamps, ...) in einen Frame geladen

## FLAC: Modellierung

- **Verbatim**
  - Das Audiosignal wird ohne Kompression kopiert
- **Linear**
  - Das Signal wird mit einer Linearfunktion angenähert
  - Stille, konstante Töne
- **Polynomfunktion**
  - Das Signal wird mittels einer Polynomfunktionen angenähert
  - Schnell, weniger genau als LPC
- **Linear Predictive Coding**
  - Vereinfachtes Modell des Klangerzeugers, z.B. des menschlichen Stimmtraktes
    - \* Summer am Ende eines Rohres
    - \* Dazu Zisch- und Knalllaute
  - Danach müssen nur die Parameter dieses Modells bestimmt (und übertragen) werden

## Frequenzfilter

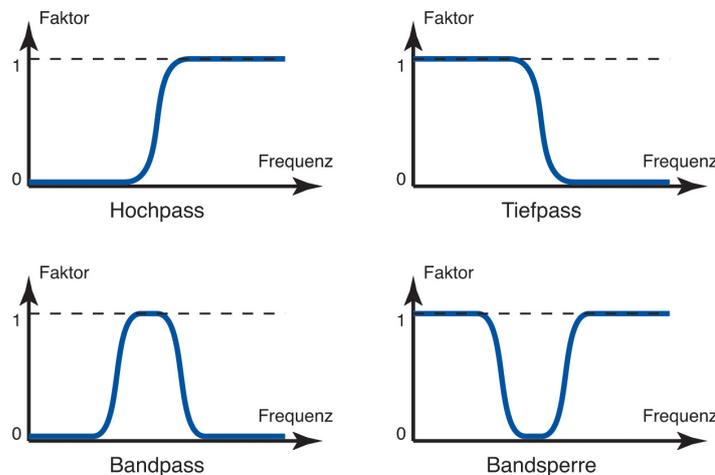


Abbildung 4.17: Die vier grundlegenden Arten von Frequenzfiltern

## Dynamik

## Dynamikkompressions

- Dynamik lässt sich als Verhältnis von größten und kleinsten Amplituden innerhalb eines Zeitfensters quantifizieren
- Je nach Länge des Fensters unterscheidet man Mikrodynamik (Sekundenbruchteile) und Makrodynamik (Sekunden und Minuten)
- Mikrodynamik macht z.B. den Charakter eines Instrumentes aus, Makrodynamik den Lautstärkeaufbau eines Musikstücks

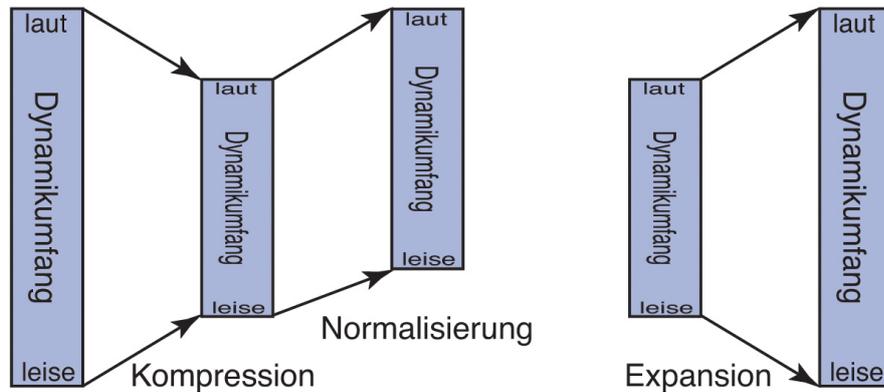


Abbildung 4.19: Kompression und Expansion der Dynamik eines Audiosignals

## Übung: Reparatur Audiospuren

- Wir stellen ihnen eine Audiodatei zur Verfügung
- Verringern Sie mit geeigneten Werkzeugen die Störgeräusche
  - Knistern, Hintergrundton
- Zur Wahl stehen z.B. die Filter oder der Equalizer
- Bringen Sie die gesprochenen Anteile auf eine gleichmäßige Lautstärke
- Schneiden Sie die Sprecherin oder den Sprecher heraus
- Schieben Sie die Spuren (falls nicht automatisch passiert) mit dem Verschiebewerkzeug wieder zusammen, sodass die Schnitte nicht mehr zu hören sind

## 7 Bilder

### Licht

- Fotografieren: Mit Licht schreiben
- Licht kann heller oder dunkler sein, unterschiedliche Farben haben
- Vereinfachte Annahme: Keine Photonen, keine Relativität
- Licht ist eine **elektromagnetische Welle**
- Charakteristische Eigenschaften
  - Ausbreitungsrichtung, Frequenz/Wellenlänge, Amplitude, Polarisation

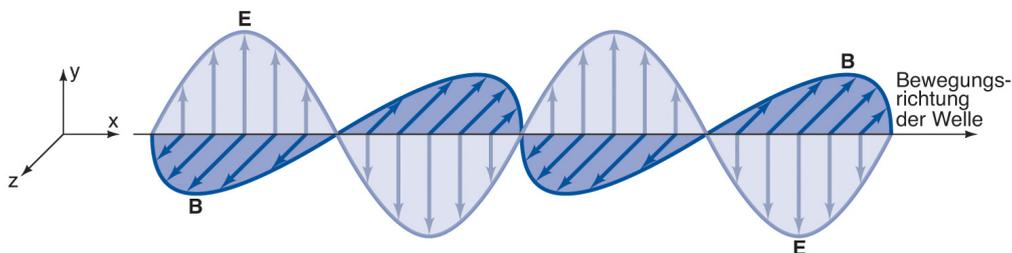
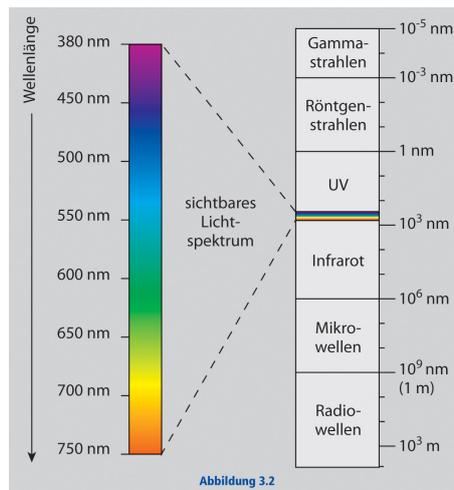


Abbildung 3.1: Zwei Wellen E und B, die sich in x-Richtung ausbreiten

## Grundlegende Größen

- **Amplitude:** Maximaler Wert in einer Schwingung
- **Phase:** Ein Durchlauf
- **Phasenverschiebung:** Verschiebung des Nulldurchgangs (0-360 Grad)
- **Frequenz:** Anzahl der Schwingungen pro Zeiteinheit
- **Wellenlänge:** Der Weg, den das Signal bei gegebener Ausbreitungsgeschwindigkeit zurücklegt
- **Polarisation:** Ebene, in der die Schwingung verläuft
- **Longitudinalwelle** vs. **Transversalwelle:** Schwingung in der oder senkrecht zur Ausbreitungsrichtung

## Frequenz



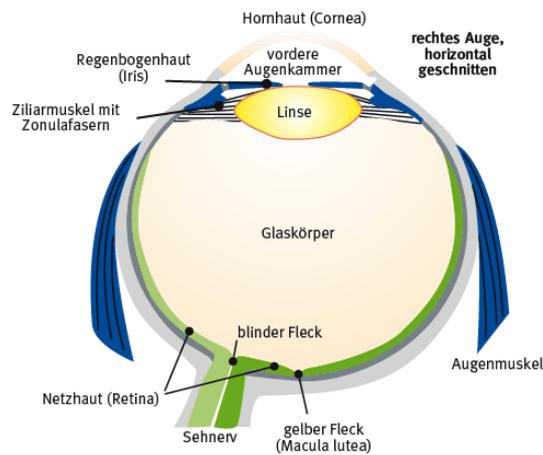
## Das menschliche Auge

- Komplexes Sinnesorgan
- Über die Linse wird das Licht auf die innen liegende Netzhaut projiziert
- Etwa 120 Millionen Sehzellen
  - ca. 113 Millionen Stäbchen, die nur Helligkeit wahrnehmen
  - ca. 7 Millionen Zapfen, die jeweils auf blaue, grüne, rote Farbtöne reagieren
- In der Tierwelt: Arten mit 2-4 Arten von Zapfen
- Bei Frauen tetrachromatisches Sehen möglich, allerdings nur ein Fall bekannt
- Ca. 8% der Männer, 1% der Frauen sind farbenblind
  - Verschiedene Formen
  - Meist können weniger Farben unterschieden werden
  - Häufigste Form: Rot-Grün-Schwäche

## Wichtige Bestandteile

- **Iris:** Reguliert die Menge des einfallenden Lichts
- **Linse:** Fokussiert ein kleines Abbild der Realität auf die Netzhaut
- **Netzhaut:** ist mit Lichtsinneszellen bedeckt die optische Reize (Stäbchen: Helligkeit, Zapfen: Farbe) wahrnehmen.
- **Gelber Fleck:** größte Dichte an Farbrezeptoren (17° des Sichtfeldes)
- **Fovea Centralis:** Bereich des schärfsten Sehens in der Mitte des Gelben Flecks (1.5-2° des Sichtfeldes)
- **Blinder Fleck:** hier verlässt der Sehnerv das Auge

## Schema



Bildquelle: [www.tg8.eu](http://www.tg8.eu)

## Gesichtsfeld

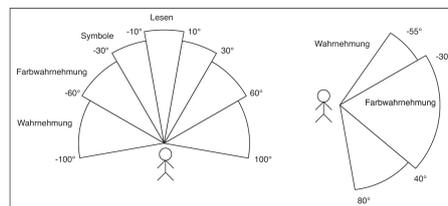
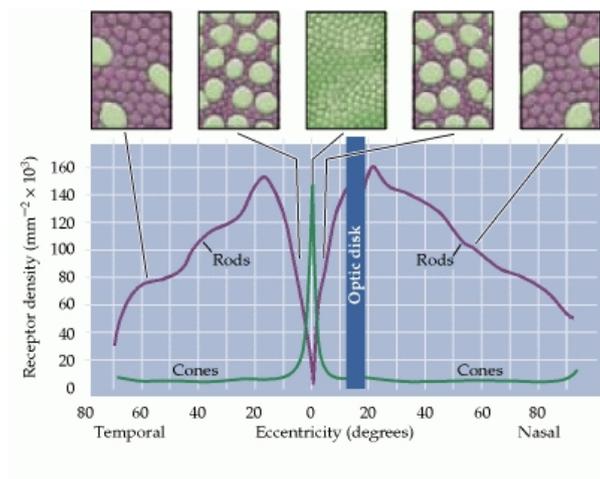


Abbildung 1.1: Sehfeld beim Menschen. Links horizontal, rechts vertikal (nach Herzog, 1994)

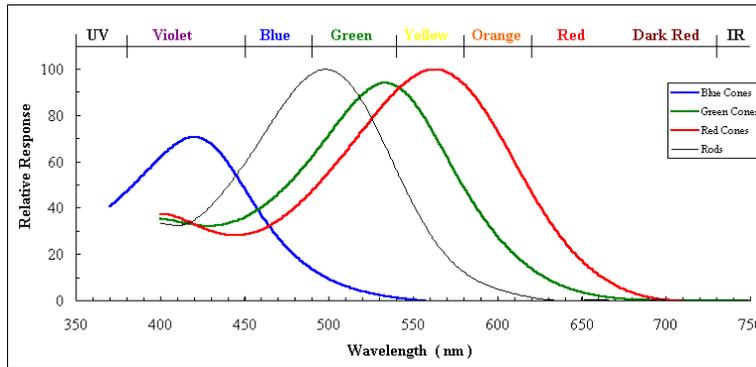
- Höchste Auflösung in der Mitte des Sehfeldes (ca 2 Grad)
  - Etwa 1 Bogenminute (1/60 Grad)
  - Ein Objekt von 3 mm Größe in ca 10 m Entfernung
- Dort finden sich viele Zapfen, aber kaum Stäbchen
- Nachts sind wir im Zentrum des Sehfeldes faktisch blind
- In der Peripherie ist das Sehen stark eingeschränkt, bis 40 Grad Augenbewegung, sonst Kopf drehen

## Verteilung Zapfen & Stäbchen



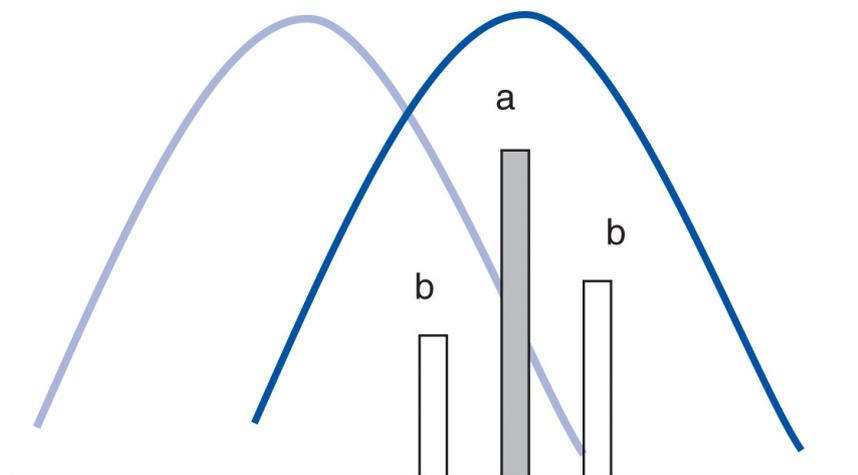
Purves et al. [2001]

## Empfindlichkeit Rezeptoren



www.unm.edu/~toolson/human\_cone\_response.htm

## Auflösung

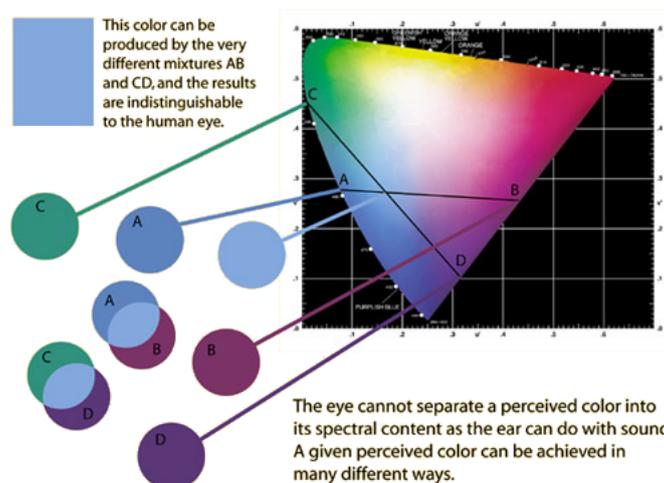


**Abbildung 3.4:** Zwei Signale (a) und (b) führen zu gleicher Farbwahrnehmung.

mögliche Farbeindruck beim Menschen kann durch RGB-Signale erzeugt werden

Umgekehrt: "Jeder"

## Erzeugung von Farbeindrücken



hyperphysics.phy-astr.gsu.edu/hbase/vision/colper.html

## Additive Farbmischung

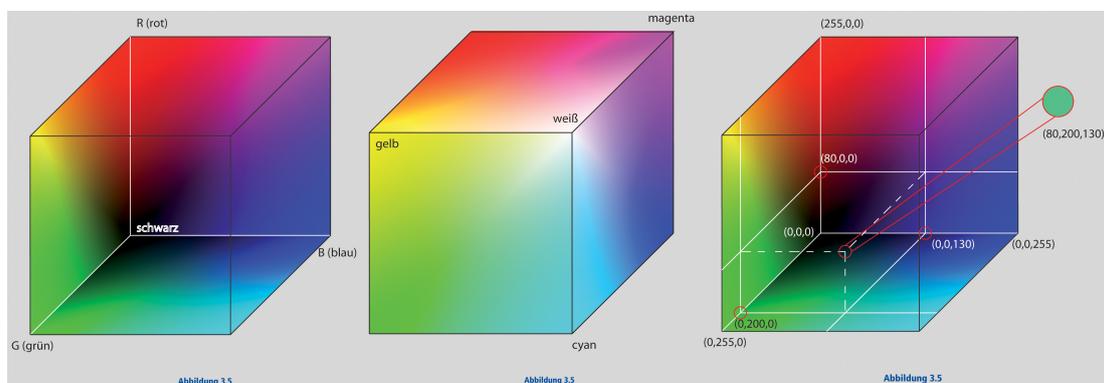
- Wir können die wahrnehmbaren Farben als Mischung von Rot, Grün und Blau herstellen

- Unser Weiß ist Rot + Grün + Blau
- **Additive** Farbmischung:
- Durch Hinzufügen von Farben zu Schwarz werden neue Farben erzeugt
  - Rot + Grün = Gelb
  - Rot + Blau = Magenta
  - Grün + Blau = Cyan
- Farbmischung bspw. bei Monitoren und Projektoren
- **Farbraum:** Mischbare Farben

### Subtraktive Farbmischung

- Farbeindruck auf Papier entsteht durch Filterung von Farbanteilen
- Grün ist die Abwesenheit von Blau und Rot
  - Blau und Rot werden absorbiert
- Mischung auf Papier heißt somit: mehr Filterung
- **Subtraktive** Farbmischung
- Im RGB-Farbwürfel werden jeweils die gegenüberliegenden Ecken gefiltert
  - Cyan: Rot wird gefiltert
  - Gelb: Blau wird gefiltert
- CMY-Modell: Cyan Magenta Yellow
- CMYK: Hinzunahme von Schwarz
  - Druckerfarbe für besseres Schwarz und geringere Kosten

### RGB-Würfel



### Weitere Farbmodelle

- Im wesentlichen zwei Gruppen:
  - Varianten von RGB/CMY(K) (Grundfarbenorientierte Modelle)
  - Luminanz-Chrominanz (Wahrnehmungsorientierte Modelle)
- Letztere arbeiten nicht mit Grundfarben, sondern trennen **Farbinformation** von **Helligkeitsinformation**
- Analog menschliche Wahrnehmung: Integration über die Antwort der Zapfen
- Vorteil: Anpassung an unterschiedliche Helligkeiten
- Bei RGB und CMY durch Änderung eines Farbwertes auch Änderung von Helligkeit und Sättigung
- Auch: Schwarzweißbild + Farbinformationen
- Umrechnung zum Teil aufwendig und nicht linear

## YCbCr

- Einfaches lineares Modell mit Helligkeitskanal: YCbCr
  - Y: Helligkeit, hoher Anteil von Grün, geringer Rot und Blau
  - Cb: Blauanteil
  - Cr: Rotanteil
- Cr und Cb viel “flächiger”
- benachbarte Bildpunkte haben oft den gleichen Wert
- zudem sind diese oft im mittleren Wertebereich
- Das Phänomen, das mehr Information in den Helligkeitswerten spricht, tritt bei vielen natürlichen Bildern auf – Farben sind gleichmäßiger verteilt als Helligkeiten
  - einheitliche Färbung bei dreidimensionaler Struktur
  - Licht und Schatten liefern Kontrastunterschiede
- Unser Sehsystem hat sich mit darauf eingestellt, es verarbeitet Helligkeitsinformationen besser
- Besonders deutlich nachts, wenn wir ohne Farben sehen
- Farbkanäle haben weniger Informationen, unser Sehsystem ist daran angepaßt



Original



RGB



YCbCr



Original

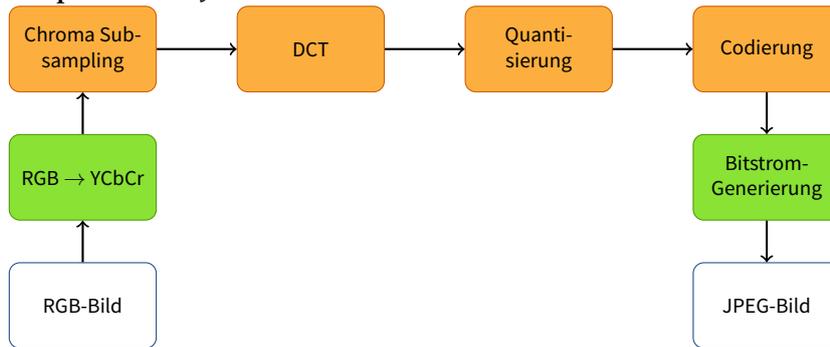


RGB



YCbCr

## Prinzipschaltbild JPEG



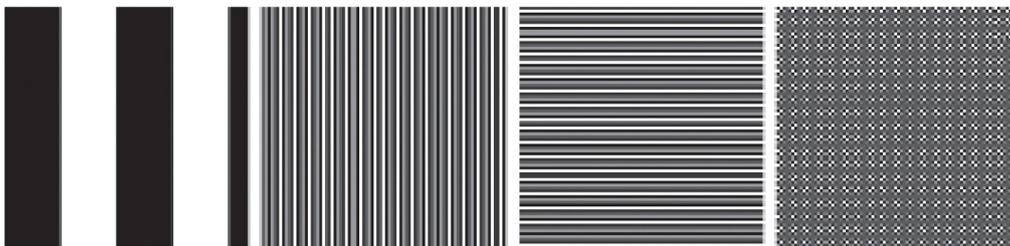
Vereinfachtes Prinzipschaltbild

## Chroma-Subsampling

- Verlustbehafteter Kompressionsschritt
- Abtastrate eines Teils des Bildes wird reduziert
- Idee: Flächigkeit der Farben im YCbCr
- Verschiedene Varianten
  - Block aus 4 Pixeln zusammenfassen
  - Zwei benachbarte Pixel zusammenfassen (horizontal oder vertikal)
- Reduziert Cr und Cb auf die Hälfte/ein Viertel
- Y-Kanal unverändert
- Subsampling nur auf den Farbkanälen
- Auswählbar: Gar nicht/Halbierung/Viertelung

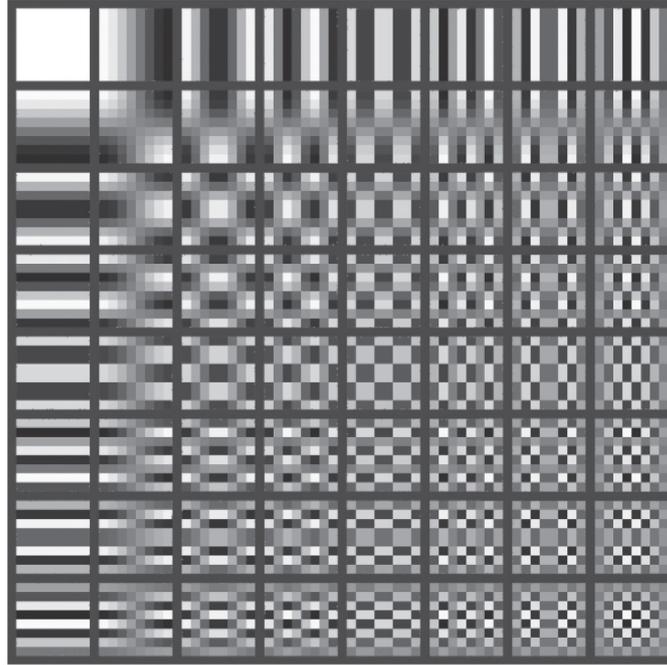
## Umcodierung in den Frequenzraum

- Folgende Schritte für die drei Kanäle getrennt
- Y, Cr, Cb unabhängig voneinander untersucht und komprimiert
- Grauwertbilder mit jeweils einem Byte pro Pixel (255 Stufen)
- Umcodiert, um danach die Quantisierung (Genauigkeit der Abtastung) zu vergrößern
- Im Frequenzraum, um die in einem Bild vorkommenden Frequenzen zu betrachten



**Abbildung 3.9:** Bilder mit unterschiedlichen horizontalen und vertikalen Frequenzanteilen, von links nach rechts: (1) niedrige horizontale Frequenzen, (2) nur hohe horizontale Frequenzen, (3) nur hohe vertikale Frequenzen, (4) Mischung hoher horizontaler und vertikaler Frequenzen

## Basismuster

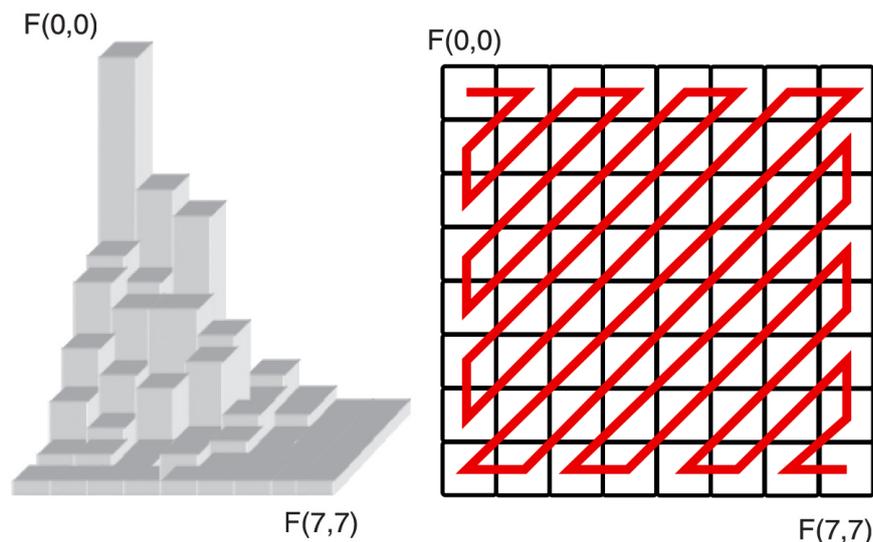


**Abbildung 3.10:** 64 Basismuster für die DCT, jedes Basismuster besteht aus 8x8 Werten, die unterschiedliche Frequenzanteile des Bildes repräsentieren. Ganz oben links ist die DC-Komponente. Die Muster weiter rechts entsprechen höheren horizontalen, die weiter unten höher werdenden vertikalen Frequenzen.

## Kompression im Frequenzraum

- Annahmen
  1. Bei vielen Bildern sind hohe Frequenzanteile seltener als niedrige
  2. Änderungen bei hohen Frequenzanteilen fallen bei der Wahrnehmung weniger ins Gewicht
- Vorgehen
  - Werte nahe Null auf Null setzen
  - Höhere Frequenzen mit weniger Bits kodieren (Auflösung reduzieren)
    - \* Quantisierung (Auflösung) wird vergrößert
    - \* Nur auf weniger gut wahrgenommenen (hohen) Frequenzanteilen
  - Daher Transformation in den Frequenzraum

## DCT Koeffizienten



**Abbildung 3.11:** DCT-Koeffizienten. Links: typische Werteverteilung bei Bildern. Der DC-Anteil  $F(0,0)$  ist hoch. Hohe Frequenzanteile (vorn und rechts) sind niedrig und oft null. Rechts: Zick-Zack-Scan zum Auslesen der AC-Koeffizienten.

### Re-Quantisierung

- Die neue Quantisierung der DCT-Koeffizienten wird mit einer Quantisierungsmatrix durchgeführt
- Dabei gibt es für jeden der 8x8 DCT-Koeffizienten einen Wert  $Q(u, v)$ , durch den dieser geteilt wird

$$F'(u, v) = \text{Round} \left( \frac{F(u, v)}{Q(u, v)} \right)$$

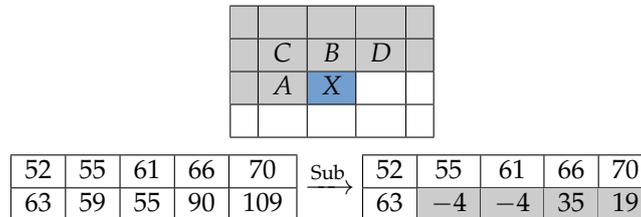
- Rundung reduziert Wertebereich für Koeffizient  $F(u, v)$
- Wesentlicher Schritt, der auch die Güte bestimmt
  - Werte  $Q(u, v)$  bestimmen, wie viele Informationen verloren gehen
- Für die Rekonstruktion muß die verwendete Quantisierungsmatrix bekannt sein

### Speicherung

- Subsampling und re-Quantisierung sind die verlustbehafteten Schritte
- Bei Speicherung ergänzt um verlustfreie Verfahren, DC-Komponente anders als andere
- Kodierung DC
  - Da benachbarte Blöcke oft ähnliche Grundtöne haben kann prädiktive Kodierung helfen – von einem Startwert werden nur Differenzen gespeichert
  - weniger Bits für kleinere Differenzwerte nötig – Huffman, andere entropiebasierte Codierung
  - DC-Komponenten ergeben ein vergrößertes Vorschau-Bild
- Kodierung AC
  - Zick-Zack-Scan, zuerst niedrige, dann hohe Frequenzen
  - Am Ende oft viele Nullen – Lauflängenkodierung
  - Danach Huffman (oder arithmetisch)
  - Standardverteilungen für Huffman vorhanden

## PNG: Grundsätze

- Zweistufiges Verfahren
  - Filter/Prediction
  - Kompression
- In der Regel hängt Farb-/Helligkeitswert eines Pixels von den Pixeln in der Nachbarschaft ab
- Differenz-/Deltakodierung:  $[1, 2, 3, 4, 5, 6] \rightarrow [1, 1, 1, 1, 1, 1]$
- Bei PNG wird jedes Pixel in Relation zu den Pixeln links, oben und links oben bestimmt



## PNG Filter

- Fünf unterschiedliche Filter zur Auswahl
  - None
    - \* Keine Vorfilterung
    - \* Es wird auf den originalen Pixeldaten gearbeitet
  - Sub
    - \* Differenzen zu dem jeweils links benachbarten Pixel
  - Up
    - \* Differenzen zu dem jeweils darüber liegenden Pixel
  - Average
    - \* Mittelwert aus darüber liegenden und dem links benachbarten Pixel
  - Paeth
    - \* Es wird aus dem links benachbarten, dem darüber liegenden und dem schräg links oben benachbarten Pixel ein sogenannter Paeth-Predictor-Wert berechnet, welcher das Pixel für die Differenzbildung bestimmt
    - \*  $A, B$  oder  $C$ , was am nächsten  $p = A + B - C$  ist
    - \* tie-breaker  $A - B - C$
- Pro Zeile & Kanal kann der beste ausgewählt werden

## Histogramm (-ausgleich)

- Bei der Bildbearbeitung Histogramme häufig wichtig
- Geben an, welche Werte wie oft in den Pixeln vorkommen
- Bei einem Grauwertbild mit 256 möglichen Werten haben wir einen Vektor  $h(w), w = 0, \dots, 255$  wobei  $h(w)$  die Anzahl der Pixel mit dem entsprechenden Wert angibt
- Typischerweise soll ein Bild den gesamten Bereich eines Histogramms nutzen, sonst wirkt es flach und besitzt wenig Kontrast
- Nutzt ein Grauwertbild mit 8 Bit pro Pixel nur die Grauwerte 50 bis 150, kann man die Werte so spreizen daß der niedrigste bei 0 und der höchste bei 255 liegt:

$$f'(x, y) = \frac{f(x, y) - h_{\min}}{h_{\max} - h_{\min}} w_{\max}$$

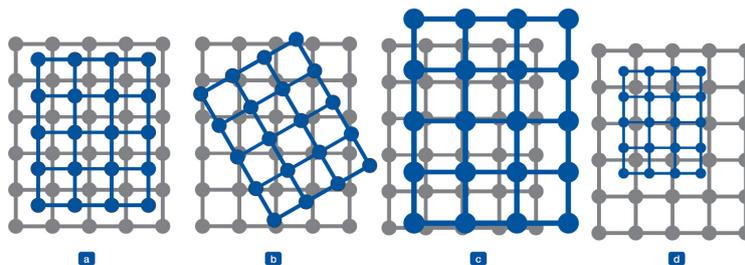
## Filter

- Im Gegensatz zu Punkt-Operationen werden bei Filtern neue Pixel berechnet, indem Informationen mehrerer Pixel des Originals betrachtet werden
- Meist in der Nachbarschaft
- Einfach: Glättung, Weichzeichner, bei dem jedes Pixel aus dem Mittelwert der Nachbarschaft besteht
- Neben den gezeigten einfachen Filtern sind auch andere denkbar
  - Wahl der Funktion, mit der die Nachbarpixel einbezogen werden



**Abbildung 3.13:** Linkes Bild: Original, Mitte: einfacher Weichzeichner, rechts: einfacher Scharfzeichner (Boost-Filter)

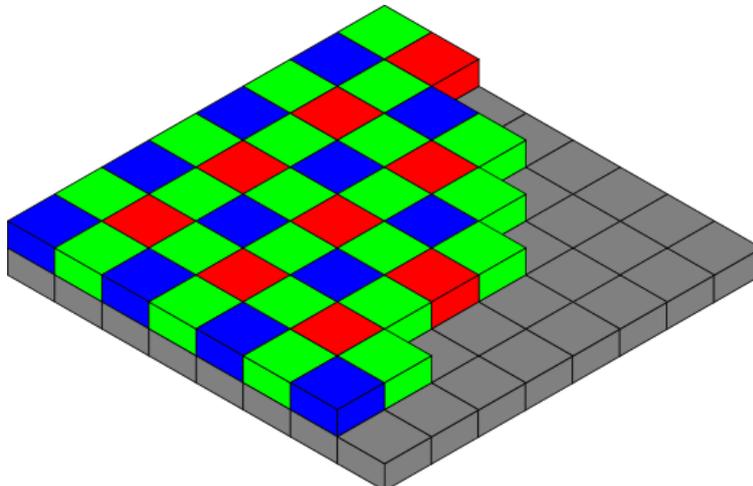
## Geometrische Operationen (contd)



**Abbildung 3.14:** Probleme bei geometrischen Operationen. Blaues Raster: transformiertes Pixelraster des Originalbildes; graues Raster: Pixelraster des Zielbildes bei (a) Verschiebung, (b) Drehung, (c) Vergrößerung, (d) Verkleinerung

- Problem: Unterabtastung, möglicher Moiré-Effekt bei Verkleinerung
  - Weichzeichner vor Verkleinerung
- Problem: Treppeneffekt bei Vergrößerung
  - Weichzeichner nach Vergrößerung

## Bayer-Filter



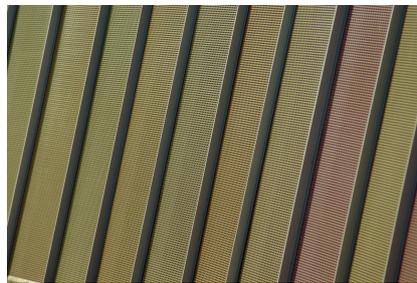
Wikipedia-User Cburnett

## Demosaicing

- Über den einzelnen lichtempfindlichen Sensoren (Pixel) liegen Farbfilter
  - Verschiedene Muster, Bayer-Filter das gebräuchlichste
- Die einzelnen Pixel enthalten jeweils nur Informationen für einen Farbkanal
- Das fertige Bild soll für jedes Pixel Informationen für alle Farben enthalten
- Lösung: Interpolation der nicht vorhandenen Farbwerte
  - Einfachster Ansatz: Interpolation aus umliegenden Pixeln
  - Komplexer: Annahmen über Bilder, z.B. Kantenerkennung
- Dieser Prozeß heißt Demosaicing und ist normalerweise Teil des Bildprozessors der Kamera, kann aber auch bei RAW-Aufnahmen im Nachhinein durchgeführt werden

## Übung: Bildbearbeitung

- Wir stellen Ihnen in den Übungen einige Bilder zur Verfügung, mit denen Sie grundlegende Funktionen der Bildbearbeitung erkunden konnten.



## 8 Video

### Bewegungswahrnehmung

- Grundsätzlich in zwei Kategorien einteilbar
  - Bewegungswahrnehmung von Objekten in Welt (Objektbewegung)
  - Bewegungswahrnehmung von sich selbst in Welt (Wahrnehmung von Eigenbewegung)
- Visuelle Objektbewegungen
  - einfache Translationen: individuelle Objekte bewegen sich im Sichtfeld
  - komplexe Bewegungen: komplexe Strukturen bewegen sich, wobei nicht Einzelbewegungen wahrgenommen werden, aber "genereller Fluss"
  - Scheinbewegungen: Bezeichnung für Bewegungseindrücke, denen keine physikalische Bewegung von Objekten zugrunde liegt

### Objektbewegung



## Wahrnehmung

- Eindruck von Bewegung komplexe Leistung des Gehirns
- Nicht nur statische Analyse, sondern auch Vergleich und Auswertung über die Zeit
- Natürliche Umgebung: kontinuierliche optische Reize
- Zeitliche Auflösung nicht beliebig genau, daher reicht es, eine beschränkte Anzahl von Bildern pro Sekunde zu präsentieren, um den Eindruck eines kontinuierlichen Signals zu erzeugen
- Wir erinnern uns an die kognitiven Grundlagen
  - Zeitliche Auflösung des visuellen Systems etwa 50ms  
(Card, Newell & Moran (1983):  $T_p \sim 100\text{ms}$  [50-200 ms])
- Aber: Von diversen Einflussgrößen abhängig
  - Beleuchtungsstärke, Leuchtdichte und Kontrast sowie Größe des Reizes und retinaler Darbietungs-ort

## Wahrnehmung (contd.)

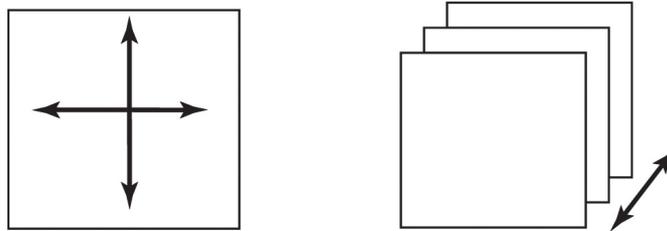
- Durch die kognitiven Leistungen des Gehirns werden bereits etwa 5 Bilder pro Sekunde zu einem Bewegungseindruck zusammengesetzt, darunter geht allerdings der Bewegungseindruck verloren
- Psychologische Grenze: 25 – 30 Bilder/Sekunde (Hz)
  - Darunter störendes Flimmern
    - \* Besonders bei Hell-Dunkel-Wechseln
    - \* Durch Reaktionszeit der Sehzellen begründet
- Flimmern verschwindet bei höheren Frequenzen
- Physiologische Grenze: 50 – 60 Bilder/Sekunde (Hz)
  - Bei etwa 100 Hz sind die Bilder nicht mehr von einer kontinuierlichen Abfolge unterscheidbar
  - Steigerung über 100 Hz nicht wahrnehmbar
- Einzelbilder nicht scharf, wir "erwarten" eine gewisse Bewegungsunschärfe

## Auge

- Weitere Effekte spielen ebenfalls eine Rolle
- Nachführen des Sehzentrums, evtl. Kopfbewegung
- Passiert im Alltag ständig
- Bei großen Leinwänden mehr Bewegung notwendig
- Daher taucht der Kinobesucher stärker in das Filmgeschehen ein (**Immersion**)
- Videobrillen helfen einerseits bei der Immersion, da nur das Videobild gesehen wird
- Andererseits fehlen bei "traditionellen" Videobrillen die Kopfbewegungen
- Lösung: Brillen, welche je nach Kopfbewegung unterschiedliche Teile des Bildes zeigen
- Die Immersion wird ebenfalls durch räumliche Aspekte gesteigert (3D)

## Kompression

- Folgende Dimensionen bieten Ansätze zur Kompression:
  - Die zwei Bilddimensionen
  - Pixeleigenschaften (Helligkeit, Farbe)
  - Zeitliche Dimension
- Mischung verlustbehafteter und verlustfreier Schritte
- Intra-/Inter-Frame Kodierung



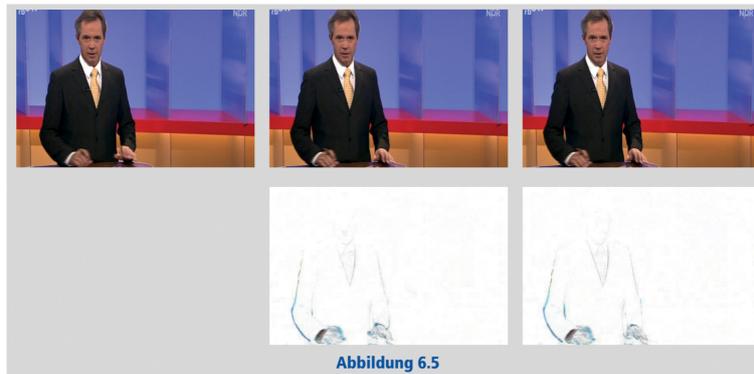
**Abbildung 6.4:** Bei der Intra-Frame-Codierung werden Redundanzen innerhalb von Einzelbildern gesucht. Bei der Inter-Frame-Codierung werden Redundanzen zwischen aufeinanderfolgenden Bildern gesucht.

### Intra-Frame Kodierung

- **Verfahren** wie in **Bildkompression** kennengelernt
- **Vektorquantisierung**
  - Aufteilung in (z.B.  $4 \times 4$ ) Pixelblöcke, Suche ähnlicher Blöcke, Übertragung der Indizes
  - Vorteile:
    - \* Schnelle Dekodierung
    - \* gut bei vielen ähnlichen Blöcken
  - Nachteile:
    - \* Aufwendige Kodierung
    - \* nicht notwendigerweise besser als z.B. DCT
- **Kontourbasierte Codierung**
  - Umrisse von Objekten erkennen, diese z.B. mit Beziér-Kurven beschreiben
  - Vorteile:
    - \* Umrisse bleiben erhalten
    - \* keine Unschärfe-Artefakte an harten Kanten
  - Nachteil:
    - \* Noch nicht effizient und zuverlässig möglich

### Inter-Frame Kodierung

- Temporale oder Inter-Frame Kodierung nutzt Redundanzen zwischen zeitlich aufeinander folgenden Bildern aus
- Zwei folgende Bilder üblicherweise ähnlich (Grundlage für den Eindruck von Bewegung)



## Differenzkodierung

- **Differenzkodierung** (frame differencing) nutzt Eigenschaft relativer Konstanz
- Nach einem Startbild werden nur die Änderungen übertragen
- Da die Differenzen klein sind werden viele Pixelwerte 0
- Die Pixel, die nicht 0 sind, sind häufig klein
- Daher kann man:
  - Pixelwerte mit weniger Bit quantisieren
  - Durch Entropie-Encoding häufig vorkommende Bits effizient speichern
  - Durch Lauflängenkodierung häufig auftretende Werte effizient speichern
- Wie bei Einzelbildern kann die Kodierung im Frequenzraum stattfinden (DCT, FFT)

## Bewegungskompensation

- In vielen Videos werden das ganze Bild oder Teilbereiche in Sequenzen verschoben
- Kamerschwenk: der ganze Ausschnitt ist in folgenden Frames versetzt
- Bewegte Objekte: Die Objekte "wandern" über den Bildausschnitt
- Falls die verschobenen Bildteile erkannt werden können muß nur die Größe des Blocks und der Bewegungsvektor gespeichert werden (Block Matching)
- Falls sich die verschobenen Bildteile verändern müssen zusätzlich die Änderungen festgestellt und gespeichert werden (Stichwort Differenzkodierung)
- Diese Verfahren werden unter dem Begriff **Bewegungskompensation** zusammengefaßt

## Frametypen

- **I-Frame – Intra-Frame**; kodiert mit vollständigen Informationen über das zu zeigende Bild.
- **P-Frame – Predicted Frame**; kodiert mit Hilfe von Inter-Frame Techniken, bezieht sich auf vorhergehende Bilder. Dieses muß vorher dekodiert werden.
- **B-Frame – Bi-Directional Predicted Frame**; mit Hilfe von Inter-Frame Techniken kodiert, die sich auf vorhergehende und auf nachfolgende Bilder beziehen können.
- Je nach Codec können P- oder B-Frames auch als Basis für (rekursive) Vorhersage anderer Bilder dienen.



Quelle: Cisco

## Digitale Filmproduktion

- Typischerweise in drei Produktionsphasen
  - Vorproduktion
    - \* Planung und Aufstellung benötigter Ressourcen
    - \* Festlegung der Länge
    - \* Musik, Sounds festlegen
    - \* Skript und/oder Storyboard erstellen
    - \* Kameratyp und Medium festlegen
  - Produktion
    - \* Drehen
    - \* Beleuchtung
    - \* Aufnahme Audio
    - \* Erstellen Graphiken, Animation, SFX
  - Postproduktion
    - \* Schnitt, Montage, Einbindung Effekte
    - \* Codierung, Auslieferung (DVD, WWW)

## Story und Plot

Filme haben einen erzählerischen Charakter. Es wird eine Kette von Ereignissen dargestellt, die Ursachen und Wirkungen in Raum und Zeit verknüpfen.

Die **Story** in einem Film beschreibt die Ereignisse der Erzählung, die entweder explizit gezeigt werden oder die Betrachter aus dem Kontext erschließen

Der **Plot** beschreibt alles, was tatsächlich im Film sichtbar gemacht wird; also insbesondere die Teile der Story, die im Film gezeigt werden.

## Drehbuch und Storyboard

Um Story und Plot schriftlich festzuhalten gibt es eine ganze Reihe von Dokumenten. Die wichtigsten sind:

Das **Drehbuch** beschreibt: "Was wird erzählt?"

Es beinhaltet alle Dialoge, Regieanweisungen und Orte als lineare, textuelle Beschreibung

Das **Storyboard** gibt Auskunft darüber: "Wie wird es erzählt?"

Es enthält detaillierte Skizzen des fertigen Films und zeigt für jede Einstellung, was zu sehen ist und was passiert.

## Beleuchtung

- In Innenräumen arbeitet man häufig mit einer Vierpunkt- oder Dreipunktausleuchtung
  - **Führungslicht/Schlüssellicht (Keylight)**: Beleuchtet die wichtigsten Objekte etwas schräg von vorne und oben und simuliert das Sonnenlicht. Es sollte nicht direkt von vorne kommen, um den räumlichen Eindruck zu betonen.
  - **Spitzlicht/Gegenlicht**: Von hinten, erzeugt klare helle Konturen um die ausgeleuchtete Person oder da Objekt. Es trennt Vorder- und Hintergrund und schafft Raum.
  - **Aufhelllicht (Fillight)**: Mildert harte Schatten durch das Führungslicht, ist aber schwächer als dieses.
  - **Hintergrundlicht (Backlight)**: beleuchtet hinter den ausgeleuchteten Objekten dem Hintergrund

## Beispiel 4-Punkt Beleuchtung

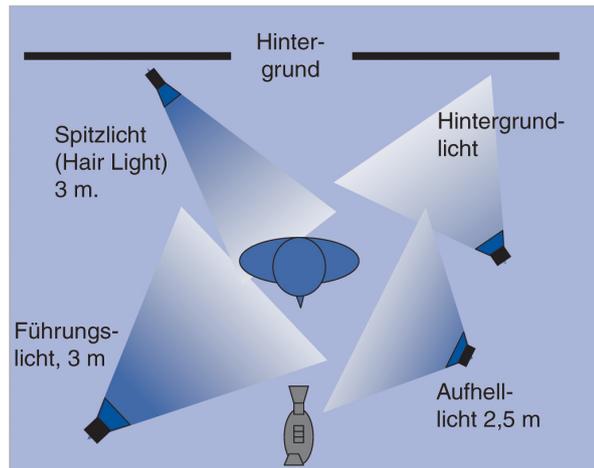


Abbildung 6.10: Aufbau einer Vierpunktausleuchtung

## Kontinuität

**Kontinuität** ist ein wichtiges Element für den Gesamteindruck des Films

- **Aktionen** dürfen nicht doppelt auftreten
- **Blickrichtung** muß beibehalten werden
- **Bewegungen** von Objekten/Personen müssen i.d.R. zwischen Einstellungen die gleiche Richtung haben.
- **Graphische Kontinuität:** ein sanfter Übergang zwischen Einstellungen (Balance der Figuren, Symmetrie, Beleuchtung, Farben)
- Der **Rhythmus** muß zur Bildeinstellung passen. Daumenregel: Einstellungen in der Totale länger als in der Halbtotalen länger als Close-Up
- **Räumliche Kontinuität**
  - Siehe auch: 180-Grad-Regel

## Audioschnitt

- Schnitte der Audiospur müssen den Schnitten der Videospur angepaßt werden
- Der Ton wird häufig genutzt, um Videosequenzen zusammen zu binden (Tonbrücke)
- Beispiele sind der J-Schnitt, der kurz vor dem Videoschnitt erscheint, oder der L-Schnitt, der danach erfolgt



Abbildung 6.12: Verzahnung von Audio- und Videospur beim J-Schnitt und beim L-Schnitt

## Beispiel Chroma Keying

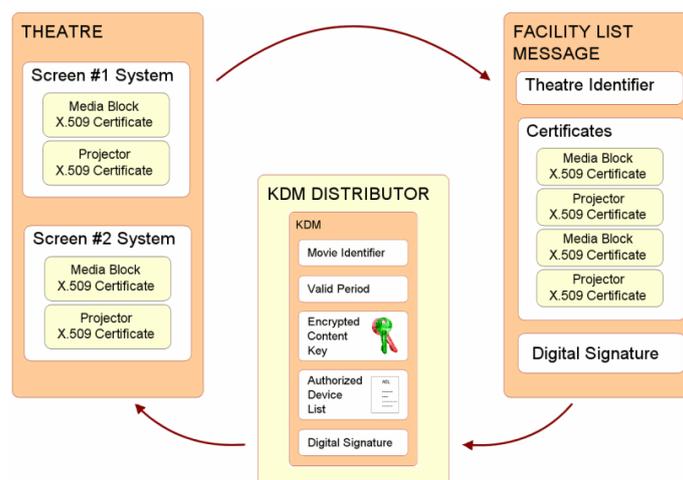


### Beispiel Depth Keying: Tiefenbild



Tiefenbild, Bildquelle: M. Rogat

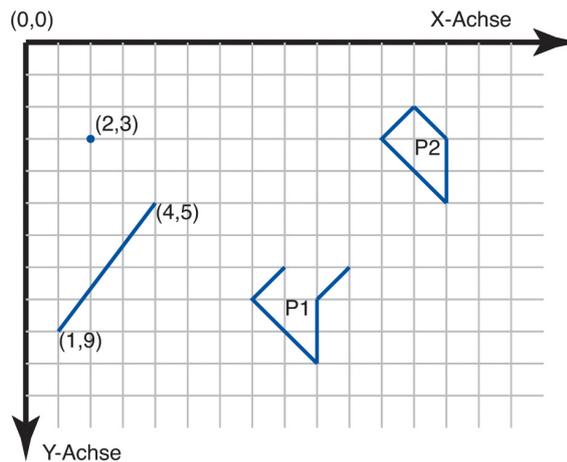
### Digitale Distribution



Quelle: MKPE Consulting

## 9 2D-Vektorgraphik

### Konzepte



**Abbildung 7.1:** Zweidimensionaler Vektorraum mit Punkt (2,3), Gerade von (1,9) nach (4,5), offenem (P1) und geschlossenem Polygon (P2)

## Kurven

- Weitere Elemente neben Punkten, Geraden, Polygonen
  - Kreis durch Mittelpunkt und Radius
  - Interpolationskurven durch Kontroll- und Stützpunkte
- Interpolationskurven:
  - Darstellung wirklich beliebiger Kurven eher unüblich, in der Praxis Annäherung
  - In der Praxis oft: Splines
  - Begriff aus dem Schiffbau: mit Hilfe von elastischen Latten, die an bestimmten Punkten fixiert harmonische (und teilweise optimale Formen) zu bilden

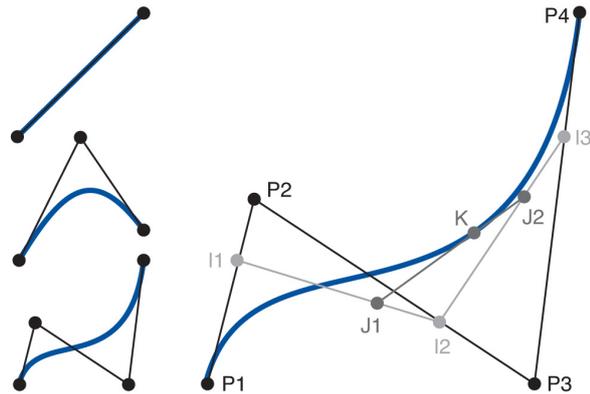
## Splines

- Ein Spline  $n$ -ten Grades ist stückweise aus Polynomen maximal  $n$ -ten Grades zusammengesetzt
- Falls die Polynome linear sind ist der Spline linear, d.h. ein Polygonzug ist ein linearer Spline
- Für jedes Segment können Randbedingungen angegeben werden:
  - Steigung
  - Krümmung
  - Krümmungsveränderung
  - Erste bis dritte Ableitung
 jeweils an den beiden Endpunkten
- Häufig in Graphikprogrammen durch Kontrolllinien angegeben
- Richtung der Kontrolllinien Steigung, Länge Steifigkeit

## Bézier-Splines

- Besondere Familie von Interpolationskurven: Bézier-Kurven
- Konstruiert aus Bernsteinpolynomen
- Aus Bézier-Kurven werden die Bézier-Splines zusammengesetzt
- Unabhängig voneinander von Pierre Bézier bei Renault und Paul de Casteljau bei Citroën entwickelt
- Nach Art der verwendeten Polynome Bézier-Kurven ersten, zweiten, dritten Grades
- Bézier-Kurven  $n$ -ten Grades beschrieben durch  $n + 1$  Kontrollpunkte
- Diese bilden das Stützpolygon
  - Bézier-Kurve ersten Grades hat ein Stützpolygon aus zwei Punkten, ist also eine Linie
- Kurvenverlauf durch den Algorithmus von Casteljau rekursiv berechenbar (siehe Graphik)

## Beispiel Bézier-Kurven



**Abbildung 7.3:** Links von oben nach unten Bézier-Kurven ersten, zweiten und dritten Grades, rechts eine Darstellung des Algorithmus von Casteljau

## Casteljau

- Um die gezeigte Kurve mit den Kontrollpunkten  $P1$  bis  $P4$  zu interpolieren werden zunächst Punkte auf den Linien des Stützpolygons interpoliert
- Die lineare Interpolation  $B$  zwischen zwei Punkten  $A$  und  $B$  ist dabei für einen Interpolationswert  $i$ :  $0 \leq i \leq 1 : B = (1 - i) * A + i * C$
- Im Beispiel:  $I1$  zwischen  $P1$  und  $P2$ ,  $I2$  zwischen  $P2$  und  $P3$ ,  $I3$  zwischen  $P3$  und  $P4$
- Zwischen den Interpolierten Punkten wird weiter interpoliert
- $J1$  wandert von  $I1$  nach  $I2$ ,  $J2$  von  $I2$  nach  $I3$
- Zwischen  $J1$  und  $J2$  wird dann der Kurvenpunkt  $K$  interpoliert
- So kann für jeden Wert von  $i$  zwischen 0 und 1 der Kurvenwert  $K$  berechnet werden

## Harmonische Spline-Kurven

- Aus den beschriebenen Segmenten lassen sich beliebig lange harmonisch gebogenen Spline-Kurven zusammensetzen
- An den Grenzpunkten durch Manipulation der Kontrollpunkte sicherstellen, daß Steigung und Krümmung benachbarter Elemente gleich sind
- Parallele Kontrolllinien = Steigung der Kurvenstücke ist identisch
- Gleich lange Kontrolllinien = Krümmung identisch
- Solcherart gebildete Bézier-Splines ahmen das Verhalten der Holzleisten nach

## Geometrische Transformationen

- Grundlegende lineare Transformationen sind:
  - Translation
  - Rotation
  - Skalierung
  - Scherung
- Die Repräsentation von Bézier-Splines macht diese invariant zu Rotation, Skalierung, Verschiebung

## Translation

Bei der Translation wird jeder betroffene Punkt um den gleichen Vektor  $(t_x, t_y)$  verschoben

$$\begin{pmatrix} x_{neu} \\ y_{neu} \end{pmatrix} = \begin{pmatrix} x_{alt} \\ y_{alt} \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} = \begin{pmatrix} x_{alt} + t_x \\ y_{alt} + t_y \end{pmatrix}$$

## (Uniforme) Skalierung

- Skalierung von  $(x_1, y_1)$  um  $s_x$  und  $s_y$
- Skalierung heißt uniform, wenn  $s_x = s_y$ :

$$\begin{pmatrix} x_{neu} \\ y_{neu} \end{pmatrix} = s \begin{pmatrix} x_{alt} \\ y_{alt} \end{pmatrix} = \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix} \begin{pmatrix} x_{alt} \\ y_{alt} \end{pmatrix} = \begin{pmatrix} s x_{alt} \\ s y_{alt} \end{pmatrix}$$

- Nicht-uniforme Skalierung:

$$\begin{pmatrix} x_{neu} \\ y_{neu} \end{pmatrix} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \begin{pmatrix} x_{alt} \\ y_{alt} \end{pmatrix} = \begin{pmatrix} s_x x_{alt} \\ s_y y_{alt} \end{pmatrix}$$

## Rotation und Scherung

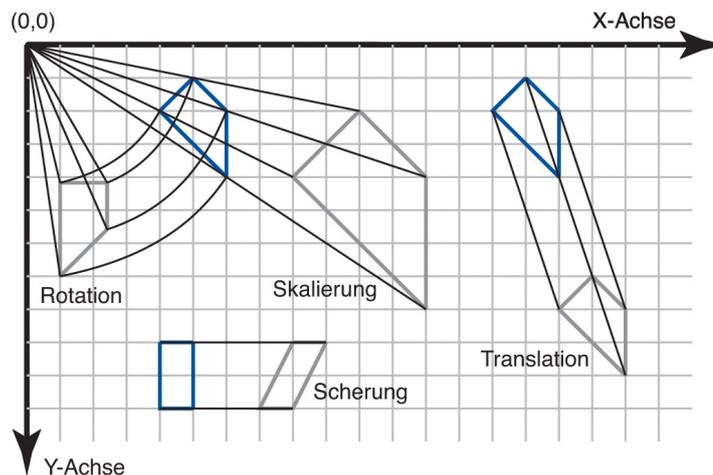
Mathematisch positive Rotation (gegen den Uhrzeigersinn) um den Winkel  $\alpha$

$$\begin{pmatrix} x_{neu} \\ y_{neu} \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} x_{alt} \\ y_{alt} \end{pmatrix} = \begin{pmatrix} \cos \alpha x_{alt} - \sin \alpha y_{alt} \\ \sin \alpha x_{alt} + \cos \alpha y_{alt} \end{pmatrix}$$

Eine Scherung entlang der  $x$ -Achse verändert  $x$ -Koordinate in Abhängigkeit von der  $y$ -Koordinate

$$\begin{pmatrix} x_{neu} \\ y_{neu} \end{pmatrix} = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{alt} \\ y_{alt} \end{pmatrix} = \begin{pmatrix} x_{alt} + m y_{alt} \\ y_{alt} \end{pmatrix}$$

## Zusammenfassendes Beispiel



**Abbildung 7.4:** Die elementaren linearen Transformationen

## Weitere Transformationen

- Mit Hilfe der definierten Transformationen sind weitere konstruierbar
- Spiegelung: Skalierung um den Faktor -1
- Rotation um andere Punkte als den Mittelpunkt durch Translation, danach Drehung, danach Translation
- Alle Operationen mit Ausnahme der Translation als Matrix-Multiplikation ausdrückbar
- Matrix-Multiplikation ist assoziativ

## Homogene Koordinaten

- Matrix-Multiplikation ist assoziativ
- Daher wäre es hilfreich, falls Translation auch so ausgedrückt werden könnte
- Lösung: Hinzunahme einer 3. Dimension bei allen Berechnungen
  - Sogenannte homogene Koordinaten, da alle Transformationen als Matrix darstellbar
  - In homogenen Koordinaten wird jeder 2D-Punkt  $(x, y)$  repräsentiert durch  $(x, y, 1)$
  - $(x, y, 1)$  und  $(x, y, W)$  repräsentieren den gleichen Punkt genau dann, wenn  $W \neq 0$
  - Matrizen erhalten in der dritten Zeile/Spalte jeweils zwei Nullen und eine 1 auf der Diagonale

## Homogene Koordinaten

$$\begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \text{ sowie } \begin{pmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{pmatrix} \Rightarrow \begin{pmatrix} m_{1,1} & m_{1,2} & 0 \\ m_{2,1} & m_{2,2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Und damit gilt für die Translation

$$\begin{pmatrix} x_{neu} \\ y_{neu} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{alt} \\ y_{alt} \\ 1 \end{pmatrix} = \begin{pmatrix} x_{alt} + t_x \\ y_{alt} + t_y \\ 1 \end{pmatrix}$$

## 2D Rendering Pipeline

- Bei Berechnung und Darstellung einer 2D-Vektorgraphik gibt es eine etablierte Abfolge von Arbeitsschritten
- Diese wird im allgemeinen als Rendering-Pipeline beschrieben

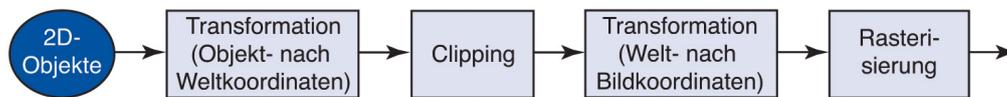
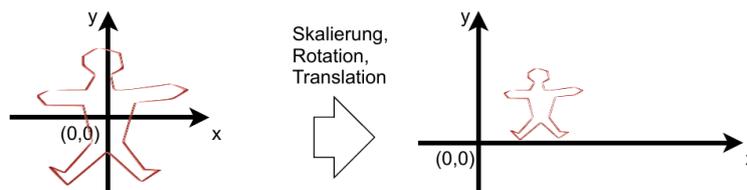


Abbildung 7.5: Die 2D Rendering Pipeline

## Objekt- und Weltkoordinaten

- 2D-Primitive werden in Objektkoordinaten beschrieben und dann in Weltkoordinaten transformiert



Quelle: Frank Steinicke

## Szenegraph

- Alle eingeführten geometrischen Primitive sind durch Punkte beschrieben
- Theoretisch könnten die Objekte alle an dem Ort, an dem sie später zu sehen sein sollen, beschrieben werden
- Praktisch meist hierarchisch organisiert
  - Gemeinsam verschieben
  - Gemeinsam animiert
  - Objekteigenschaften wie Strichbreite, Füllfarbe für Gruppen von Objekten festgelegt werden können
- Um diese effizient anwenden zu können werden die Objekte in einem Szenegraphen organisiert

## Szenegraph

- An den Blättern geometrische Objekte
- An den inneren Knoten Transformationen oder Gruppierungen
- Im einfachsten Fall ein Baum
- Bei mehrfacher Verwendung eines Objektes/einer Gruppe ein DAG (Gerichteter Azyklischer Graph)
- Mehrfachnutzung eines Objektes spart Modellierungsaufwand
- Ein Objekt kann dann durch mehrere Transformation an die jeweils richtige Stelle verschoben werden
- Geometrische Eigenschaften werden aufmultipliziert
- Nichtgeometrische Eigenschaften werden vererbt
- Die Definition, was ein Szenegraph ist, wird unterschiedlich gesehen

## Clipping

- Clipping von Punkten: Punkt liegt innerhalb des Sichtfensters
- Clipping beliebiger Polygone auf Clipping von Linien zurückführen
- Klassisches Verfahren für das Clipping gerader Linien am einem rechteckigen Ausschnitt ist das Line Clipping von Cohen und Sutherland
  - Gegeben Linie mit Start- und Endpunkten  $P_1$  und  $P_2$  sowie ein Sichtfenster mit Koordinaten  $x_{min}$ ,  $x_{max}$ ,  $y_{min}$  und  $y_{max}$
  - Unterteile den Vektorraum in 9 Teilbereiche aufgeteilt
  - Jeder Teilraum erhält einen 4 Bit langen Code

## Bitmuster Cohen und Sutherland

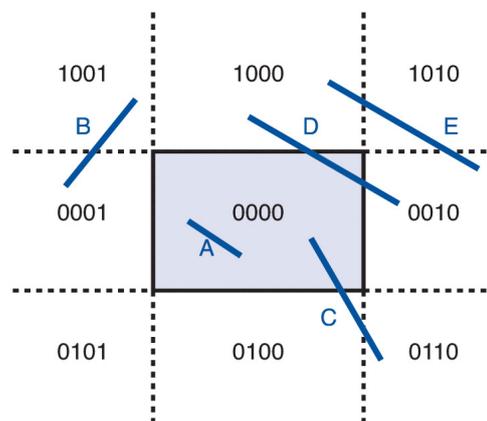


Abbildung 7.7: Line Clipping nach Cohen und Sutherland

## Cohen Sutherland (contd.)

- Start- und Endpunkt der zu schneidenden Linie wird der Code des Teilraums zugewiesen, in dem er liegt
  1. Codes  $P_1$  und  $P_2$  mit "or" verknüpft **ergeben** 0000: Objekt liegt im Viewport (Beispiel A:  $0000 \vee 0000 = 0000$ )
  2. Codes  $P_1$  und  $P_2$  mit "and" verknüpft **ergeben nicht** 0000:  $P_1$  und  $P_2$  liegen auf der gleichen "Seite" außerhalb des Viewports, das Objekt kann weggelassen werden (B:  $0001 \vee 1000 = 1001$ ;  $0001 \wedge 1000 = 0000$ )
  3. Ansonsten muss die Linie – abhängig vom Code – geschnitten werden

- Falls der Code von  $P_1$  nicht 0000 ist, muß die Gerade mit den von dort erreichbaren Rändern des Viewports geschnitten werden
  - Vom Feld 0010 ist z.B. nur der rechte Rand zu erreichen (Beispiele D und E)
- Schneidet die Linie das Randsegment wird der Schnittpunkt  $S$  als neuer Wert von  $P_1$  bestimmt
- Falls auch der Code von  $P_2$  nicht 0000 ist muß analog geschnitten werden

### Welt- nach Bildkoordinaten

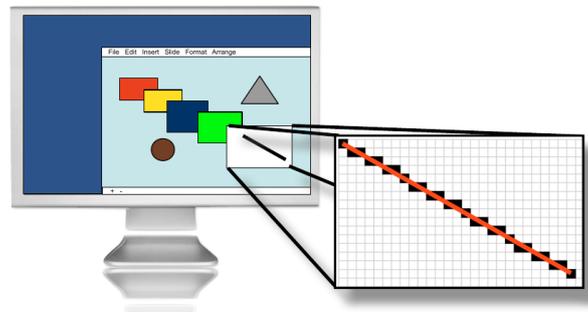
- Die darzustellenden Primitive wurden mit Hilfe des Szenegraphen an die richtige Stelle transformiert
- Die Darstellung wurde mittels Clipping auf ein definiertes Sichtfenster begrenzt
- Jetzt muß in Weltkoordinaten umgerechnet werden

$$x_{\text{bild}} = x_{0_{\text{bild}}} + (x_{\text{welt}} - x_{0_{\text{welt}}}) * (x_{1_{\text{bild}}} - x_{0_{\text{bild}}}) / (x_{1_{\text{welt}}} - x_{0_{\text{welt}}})$$

$$y_{\text{bild}} = y_{0_{\text{bild}}} + (y_{\text{welt}} - y_{0_{\text{welt}}}) * (y_{1_{\text{bild}}} - y_{0_{\text{bild}}}) / (y_{1_{\text{welt}}} - y_{0_{\text{welt}}})$$

### Rasterisierung

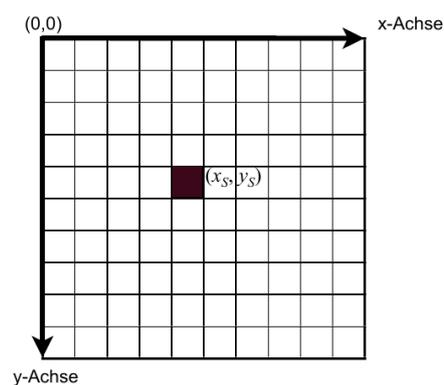
- Objekte in Bildschirmkoordinaten müssen in Pixel-Farbinformationen transformiert werden



Quelle: Frank Steinicke

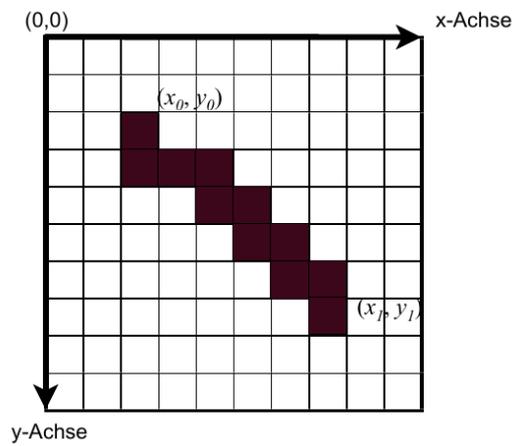
### Beispiel: Rasterisierung eines Punktes

- Farbe des Pixels, das den Punkt enthält, wird angepaßt



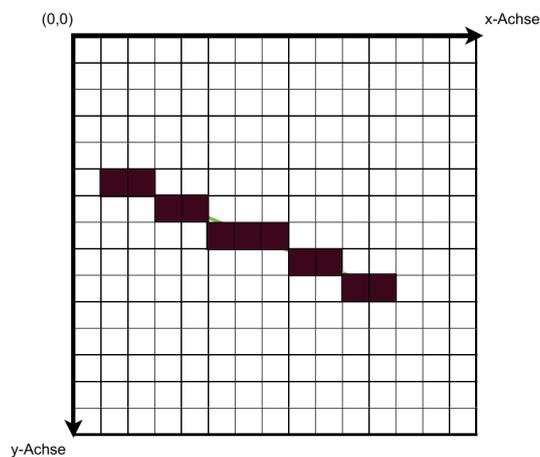
Quelle: Frank Steinicke

### Naive Rasterisierung einer Geraden



Rundung berechneter Werte  
Quelle: Frank Steinicke

### Bresenham-Rasterisierung einer Geraden



Wechsel der Änderungen  
Quelle: Frank Steinicke

### Antialiasing: Algorithmus von Wu

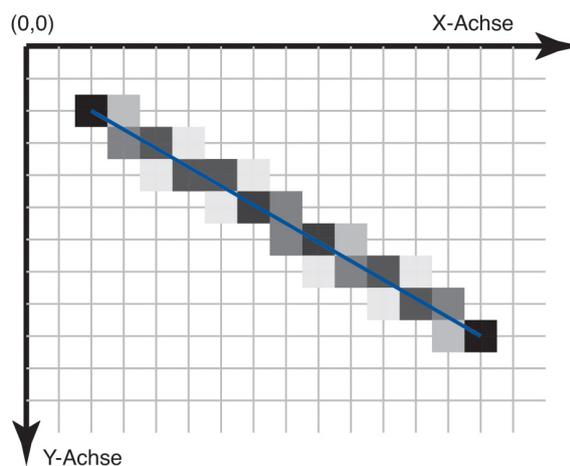
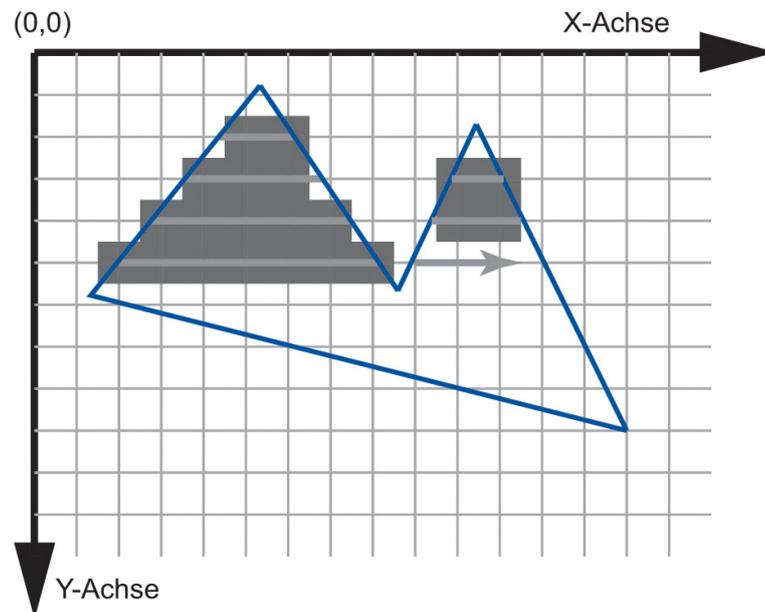


Abbildung 7.10: Rasterisierung einer Linie nach dem Algorithmus von Wu

### Scanline: Füllen von Polygonen

- Bestimme für jede Zeile von Pixeln (Scanline) die Schnittpunkte mit den Kanten des Polygons und sortiere sie aufsteigend nach der x-Koordinate

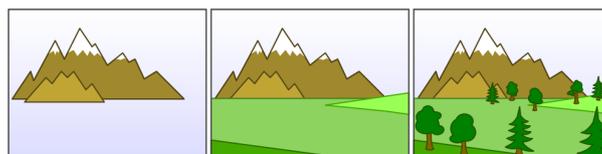
- Ermittle für jedes Pixel die Parität
- Färbe alle Pixel mit ungerader Parität mit der Füllfarbe



**Abbildung 7.12:** Ausfüllen eines Polygons mit dem Scanline-Algorithmus

### Painter's Algorithm: Verdeckungen

- Wie geht man mit mehreren, sich gegenseitig teilweise verdeckenden, Polygonen um?
- Genauere Betrachtungen gehören eher in die Generative Graphische Datenverarbeitung (Computergraphik)
- Hier: Grundidee des **Painter's algorithm**:
- Da alle vorkommenden geometrischen Formen in einer Ebene liegen genügt es zur Behandlung von Verdeckungen in 2D-Graphik alle Primitive in der richtigen Reihenfolge zu zeichnen (von hinten nach vorne)



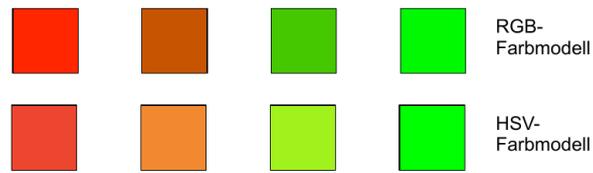
Wikipedia: Painter's algorithm

### Keyframe-Animation

- Einfachste Art, die Veränderung eines Parameters anzugeben, ist dessen Wert zu zwei Zeitpunkten anzugeben
- Z.B. Zeitpunkt 0 (Start der Animation, Laden der Datei, Beginn der Darstellung) und dann nach z.B. 10 sek
- Dazwischen wird (linear) interpoliert
- Dieses Vorgehen heißt Schlüsselbildanimation oder **Keyframe-Animation**
- Dabei können wir mehr als 2 Zeitpunkte angeben
- Je nach Software können auch andere als lineare Interpolationen zwischen einzelnen Keyframes definiert werden
  - Beschleunigen und Bremsen (ease-in, ease-out)
  - Interpolation mittels Spline

## Animation von Farben

- Animation von Farbe nicht unmittelbar klar (Farbverläufe in welchem Farbraum?)



Quelle: Frank Steinicke

## Übung: SVG-Modellierung

1. Modellieren Sie ein von der Seite gesehenes Auto in SVG, dessen Räder eine erkennbare Struktur haben (z.B. Speichen) und die sich beim Fahren drehen.
  2. Nun lassen Sie das Auto in verschiedene Richtungen über den Bildschirm fahren.
  3. Entwerfen Sie zuerst eine Transformationshierarchie.
  4. Welche Objekte müssen wie gruppiert und animiert werden, damit sich die Räder samt Speichen um ihre Mittelachse drehen, das Auto nicht ohne seine Räder losfährt und die Fahranimation nur einmal geschrieben werden muß?
- Geben Sie ihre (Zwischen-) Ergebnisse im Learnweb ab oder stellen sie diese in der Übung vor
    - Vereinfachungen möglich

## 10 3D-Graphik

### Grundlegende Elemente und Funktionen

- Zur Darstellung von 2D-Ausgaben müssen ursprünglich in 3D berechnete Darstellungen auf 2D projiziert werden
- Dazu benutzen wir eine virtuelle Kamera
- Reale Kamera hat perspektivische Projektion, virtuell sind andere Projektionen möglich
- Zur Darstellung der Objekte sind noch Oberflächeneigenschaften und Lichtquellen wichtig
- Wir haben also: 3D-Modelle, Ihre Oberflächen, Lichtquellen, Kamera
- Für eine animierte Darstellung können diese Elemente beweglich oder veränderlich sein
- Ein standardisierter Verarbeitungsprozeß, die 3D Rendering Pipeline, hat sich etabliert und wird weitgehend in Hardware unterstützt

### Koordinatensysteme

- Analog zu 2D: kartesisches Koordinatensystem mit linear gleich aufgeteilten Achsen
- Zwei Konventionen; rechtshändiges Koordinatensystem (häufiger) oder linkshändiges

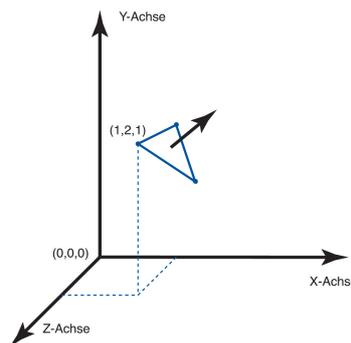


Abbildung 8.1: Rechtshändiges dreidimensionales Koordinatensystem, darin eingezeichnet ein Polygon aus drei Punkten mit seiner Flächennormalen

## Lineare Transformationen

Translation als Addition

$$\begin{pmatrix} x_{neu} \\ y_{neu} \\ z_{neu} \end{pmatrix} = \begin{pmatrix} x_{alt} \\ y_{alt} \\ z_{alt} \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} = \begin{pmatrix} x_{alt} + t_x \\ y_{alt} + t_y \\ z_{alt} + t_z \end{pmatrix}$$

Skalierung

$$\begin{pmatrix} x_{neu} \\ y_{neu} \\ z_{neu} \end{pmatrix} = \begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{pmatrix} \begin{pmatrix} x_{alt} \\ y_{alt} \\ z_{alt} \end{pmatrix} = \begin{pmatrix} s_x x_{alt} \\ s_y y_{alt} \\ s_z z_{alt} \end{pmatrix}$$

## Lineare Transformationen

Rotation als Kombination von 3 elementaren Rotationen

$$\begin{pmatrix} x_{neu} \\ y_{neu} \\ z_{neu} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} x_{alt} \\ y_{alt} \\ z_{alt} \end{pmatrix} = \begin{pmatrix} x_{alt} \\ \cos \alpha y_{alt} - \sin \alpha z_{alt} \\ \sin \alpha y_{alt} + \cos \alpha z_{alt} \end{pmatrix}$$

$$\begin{pmatrix} x_{neu} \\ y_{neu} \\ z_{neu} \end{pmatrix} = \begin{pmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{pmatrix} \begin{pmatrix} x_{alt} \\ y_{alt} \\ z_{alt} \end{pmatrix} = \begin{pmatrix} \cos \beta x_{alt} + \sin \beta z_{alt} \\ y_{alt} \\ \cos \beta z_{alt} - \sin \beta x_{alt} \end{pmatrix}$$

$$\begin{pmatrix} x_{neu} \\ y_{neu} \\ z_{neu} \end{pmatrix} = \begin{pmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{alt} \\ y_{alt} \\ z_{alt} \end{pmatrix} = \begin{pmatrix} \cos \gamma x_{alt} - \sin \gamma y_{alt} \\ \sin \gamma x_{alt} + \cos \gamma y_{alt} \\ z_{alt} \end{pmatrix}$$

## Homogenes Koordinatensystem

Homogenes Koordinatensystem

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \Rightarrow \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \begin{pmatrix} m_{1,1} & m_{1,2} & m_{1,3} \\ m_{2,1} & m_{2,2} & m_{2,3} \\ m_{3,1} & m_{3,2} & m_{3,3} \end{pmatrix} \Rightarrow \begin{pmatrix} m_{1,1} & m_{1,2} & m_{1,3} & 0 \\ m_{2,1} & m_{2,2} & m_{2,3} & 0 \\ m_{3,1} & m_{3,2} & m_{3,3} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Translation als Multiplikation

$$\begin{pmatrix} x_{neu} \\ y_{neu} \\ z_{neu} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{alt} \\ y_{alt} \\ z_{alt} \\ 1 \end{pmatrix} = \begin{pmatrix} x_{alt} + t_x \\ y_{alt} + t_y \\ z_{alt} + t_z \\ 1 \end{pmatrix}$$

## Punkte, Geraden, Polygone, Polygonnetze

- Analog zu 2D werden aus Punkten und Geraden Polygone zusammengesetzt
- Für effiziente Berechnung von Beleuchtungsmodellen ist es wichtig, daß diese planar sind, d.h. daß alle Punkte in der gleichen Ebene liegen und die Flächennormale an jeder Stelle gleich ist
- Dreiecke erfüllen dies trivialerweise, für Polygone mit mehr Seiten Überprüfung aufwendig
- Jedes beliebige planare Polygon kann aber in Dreiecke aufgeteilt werden (Triangulation)
- Daher haben sich dreieckige Polynome als elementares Flächenstück durchgesetzt
- Im Beispiel des Koordinatensystems haben wir ein solches Dreieck mit der **Flächennormalen** bereits gesehen

## Punkte, Geraden, Polygone, Polygonnetze (contd.)

- Diese Polygone können zu einem **Polygonnetz** (Polygon Mesh) zusammengesetzt
- Da sich benachbarte Dreiecke immer zwei Eckpunkte teilen lässt sich eine effektive Repräsentation finden
- Durchnummerierte Liste von Eckpunkte wird gespeichert
- Dann Liste von Dreiecken speichern, die auf diese Indizes Bezug nimmt
- Diese Datenstruktur heißt **Indexed Face Set**
- Darüber hinaus gibt es noch weitere, stark optimierte Darstellungen, auf die wir hier nicht eingehen

## Geometrische Primitive, CSG

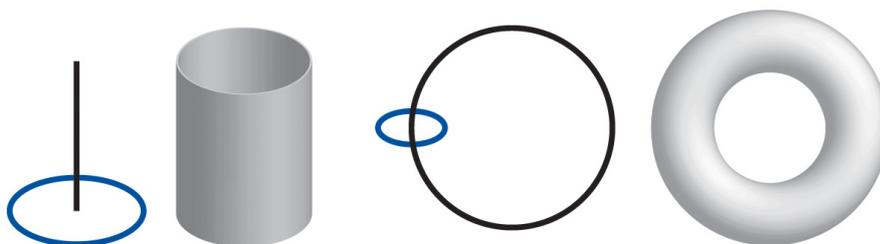
- Neben Polygonmodellen bieten 3D-Graphikprogramme (bzw. Bibliotheken) häufig die Möglichkeit, geometrische Primitive zu erzeugen
  - z.B. Quader, Zylinder, Kugel, Ring
- Aus diesen können durch boolesche Verknüpfung der Raumvolumen neue Formen erzeugt werden
- Diese Art der Konstruktion von Primitiven heißt **Constructive Solid Geometry**
- Die erzeugten Körper werden oft als **boolesche Körper** bezeichnet



**Abbildung 8.2:** Boolesche Verknüpfung eines Quaders mit einem Zylinder

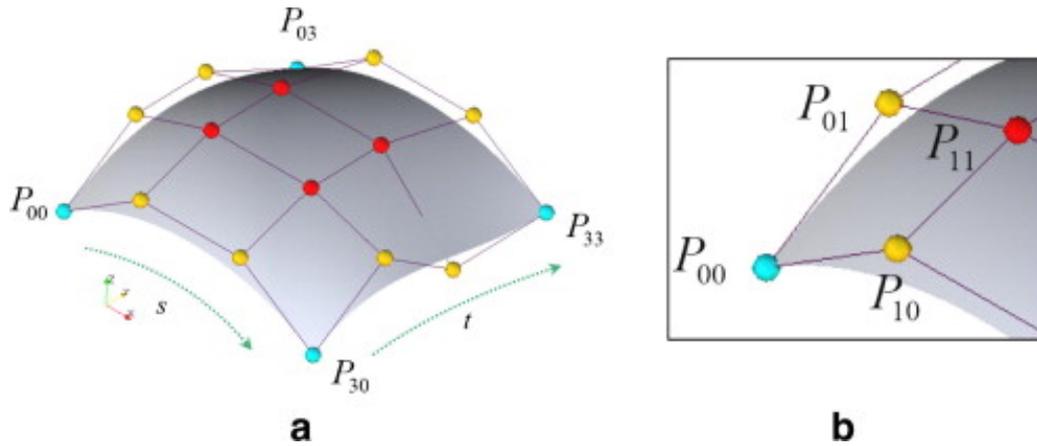
## Extrusions- und Rotationskörper

- Weitere Möglichkeit: **Extrusions-** und **Rotationskörper**
- Extrusion: Eine gegebene Grundfläche wird entlang eines beliebigen Pfades (bspw. einer Geraden) immer wiederholt
  - Aus einem Kreis entsteht ein Zylinder
- Rotation: Bei einem Rotationskörper ist dieser Pfad ein geschlossener Kreis
  - Aus einem Kreis entsteht ein Ring
- Konstruktion von Gegenständen wie Vasen oder Flasche sehr effizient



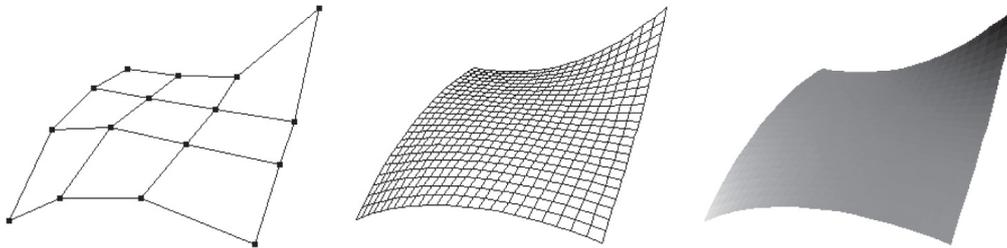
**Abbildung 8.3:** Extrusionskörper und Rotationskörper, jeweils mit zugehöriger Grundform (blau) und Pfad (schwarz), aus denen sie erstellt wurden

## Freiformflächen: Bézier Patch



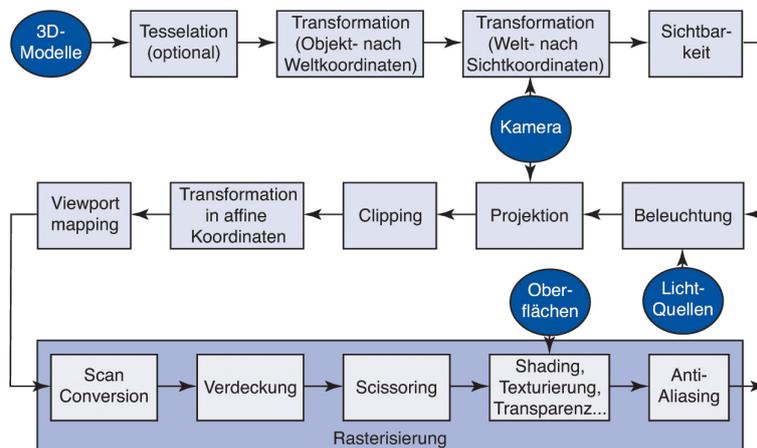
Quelle: Czarny & Huysmans, 2008

## Approximation Bézier Patch



**Abbildung 8.4:** Bikubischer Bézier-Patch: die 16 Kontrollpunkten (links), Annäherung durch Polygone (Mitte) und schattierte Darstellung (rechts)

## Rendering Pipeline



**Abbildung 8.5:** Die 3D Rendering Pipeline: der Weg vom 3D-Modell zum Bild unter Verwendung der Kamera, Lichtquellen und Oberflächenbeschreibungen

## Tesselation

- Tesella (lat.) kleines, meist rechteckiges Mosaiksteinchen
- Aus diesen Steinchen mit ebener Oberfläche lassen sich auch gekrümmte Flächen annähern
- In der CG werden gekrümmte Flächen bei der Tesselation in so kleine Polygone aufgeteilt daß sie dem Betrachter noch hinreichend gekrümmt erscheinen
- Sind die Polygone Dreiecke spricht man auch von Triangulation

## Szenegraph

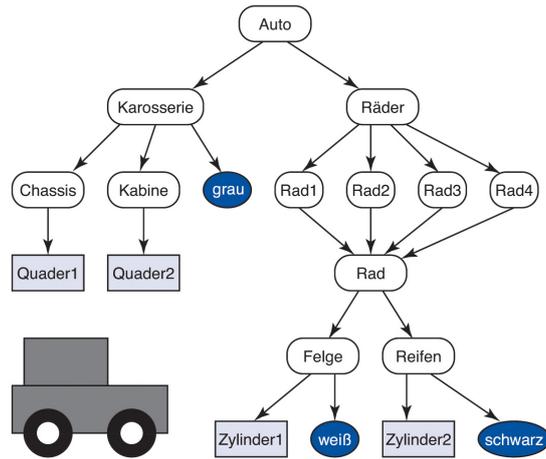
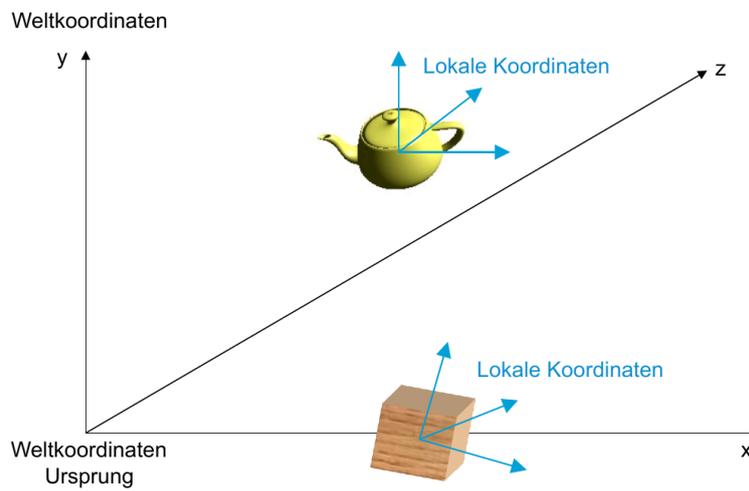


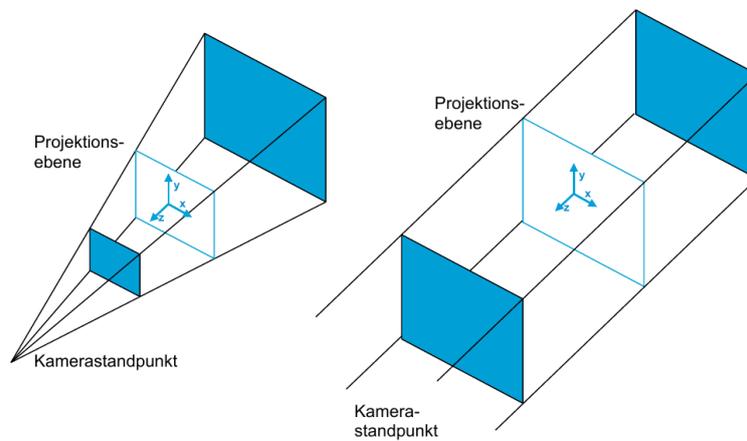
Abbildung 8.7: Szenegraph für ein Auto mit vier Rädern, in dem Geometrien, Oberflächen und ganze Baugruppen mehrfach verwendet werden

## Objekt- und Weltkoordinaten



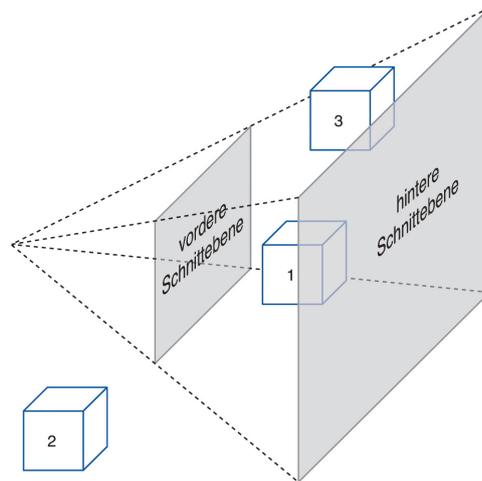
Quelle: Stefan Schlechtweg

## Kamerakoordinaten



Quelle: Stefan Schlechtweg

## Sichtvolumen



**Abbildung 8.9:** Sichtvolumen der Kamera mit vorderer und hinterer Schnittebene. Objekt 1 liegt vollständig innerhalb des Sichtvolumens, Objekt 2 vollständig außerhalb und Objekt 3 schneidet das Sichtvolumen.

## Sichtbarkeitsbestimmung

- View Frustum Culling
  - Jedes Objekt durch umschreibenden Quader (Bounding Box in 3D) annähern und mit 4 Vergleichen ermitteln, ob der Quader komplett innerhalb/außerhalb des Sichtvolumens liegt
  - Vollständig enthaltene werden später berücksichtigt
  - Vollständig außerhalb liegende können ignoriert werden
  - Für schneidende Objekte ggf. für Unterobjekte im Szenegraphen Test wiederholen, im Zweifelsfall berücksichtigen
- Back Face Culling
  - (Einseitige) Polygone, deren Normalenvektoren von der Kamera weg zeigen, aus der Pipeline entfernen
  - Grundidee: Bei geschlossenem Polygonnetz (ohne Löcher) können solche Polygone nur auf der Rückseite von Objekten vorkommen, sie sind somit durch anderen Polygone verdeckt

## Lichtquellen: Ambient

- Wir arbeiten mit verschiedenen Vereinfachungen im Vergleich mit natürlichem Licht
- Am einfachsten zu berechnende Vereinfachung ist das ambiante Licht
- In der Natur werden stets Lichtanteile von Objekten, dem Boden oder Wolken diffus reflektiert, daher ist überall etwas Licht
- Vereinfachende Modellierung, indem im Modell überall Licht von gegebener Helligkeit ohne Richtung vorhanden, die den Polygonen eine Grundhelligkeit verleiht

## Lichtquellen: Gerichtet

- Auch das gerichtete Licht wird vereinfacht modelliert
  - einerseits nimmt die Intensität quadratisch mit dem Abstand ab
  - andererseits ist die natürliche Hauptquelle Sonne so weit entfernt, daß die quadratische Abnahme praktisch nicht mehr wahrnehmbar ist
- Sonnenlicht wird daher allein durch Richtung und Helligkeit angenähert und erzeugt überall in der Szene gerichtetes Licht gleicher Intensität, das keine oder harte Schatten wirft
- Mit diesem "Sonnenlicht" und dem ambienten Licht können bei akzeptablem Rechenaufwand bereits akzeptable Beleuchtungssituationen geschaffen werden
- "Sonnenlicht" etwas wärmer, ambiantes Licht etwas blauer (sonnige Tage) oder weißer (bedeckte Tage)

## Lichtquellen: Punkt, Spot

- Lokale Lichtquellen wie Kerzen, Glühbirnen als Punktlichter
  - Beschrieben durch Position und Lichtfarbe/Intensität
  - Gleichmäßig in alle Richtungen strahlend, Intensität mit dem Quadrat der Entfernung abnehmend
  - Statt rein quadratischer Abnahme oft linearer Anteil
  - Zur Darstellung benutzte Monitore haben geringen Kontrastumfang, das "falsche", etwas kontrastarme Licht wird als natürlich angesehen
- Beschränkung des Abstrahlwinkels mit einem Kegel führt zu Spotlicht
- Die Nähe zum Rand des Lichtkegels beeinflusst die Intensität (Falloff)
  - Falloff kann hart oder weich sein, und näher oder entfernter von der Mitte des Kegels beginnen
- Sowohl Punkt wie Spotlight werfen harte Schatten

## Lichtquellen: Fläche

- Am aufwendigsten: Weich umrandete Schatten mit einer Flächenlichtquelle
- Wie 3D-Objekte durch Geometrie modelliert, die von jedem Punkt Licht aussenden
- Annäherung über verteilte Punktlichtquellen
- Aufwendig, aber beste Ergebnisse mit weich umrandeten Schatten

## Beispiel Lichtquellen

Gerichtet – Punkt – Spot



## Reflexion: Ambient

- Für ambiente Reflexion wird ein Koeffizient angegeben, der beschreibt, welcher Anteil des ambienten Lichts mit welcher Intensität gleichmäßig reflektiert wird und in den Gesamteindruck eingeht

$$I_a = k_a I_{al}$$

- mit  $I_a$  als reflektiertem Licht,  $k_a$  Reflexionskoeffizient für ambientes Licht,  $I_{al}$  Intensität des ambienten Lichts
- Kann eingesetzt werden, um völlig schwarze Schatten zu vermeiden

## Reflexion: Diffuse

- Vollständig matter Körper weist bei einer bestimmten Beleuchtung die gleiche Farbe und Helligkeit auf, egal, von wo man ihn betrachtet
- Diese Art heißt diffuse oder Lambertsche Reflexion

$$I_d = k_d I_L \langle L, N \rangle$$

- mit  $I_d$  als reflektiertem Licht,  $k_d$  Reflexionskoeffizient für diffuse Beleuchtung,  $I_L$  Intensität der betrachteten Lichtquelle und  $\langle L, N \rangle$  Skalarprodukt zwischen Flächennormale und Winkel zur Lichtquelle
- Die diffuse Reflexion nimmt mit dem Cosinus des Winkels des Auftreffens des Licht auf das Objekt ab

## Reflexion: Glanz

- Die spekulare oder Glanzreflexion folgt einem ähnlichen Gesetz
- Ein perfekter Spiegel reflektiert mit Einfallswinkel = Ausfallswinkel
- Die meisten Flächen sind keine perfekten Spiegel
- Reflexion am stärksten bei Einfallswinkel = Ausfallswinkel, aber die Intensität läuft danach nicht gleich auf 0 zurück, sondern sinkt langsam

$$I_s = k_s I_L \langle R, N \rangle^n$$

- mit  $I_s$  als reflektiertem Licht,  $k_s$  Reflexionskoeffizient für spekulare Beleuchtung,  $I_L$  Intensität der betrachteten Lichtquelle und  $\langle R, N \rangle$  Skalarprodukt zwischen Flächennormale und Reflexionswinkel
- Hierbei ergibt ein kleiner Wert für  $n$  einen weichen Lichtabfall, ein großer Wert scharfe Glanzpunkte

## Beleuchtungsmodell von Phong

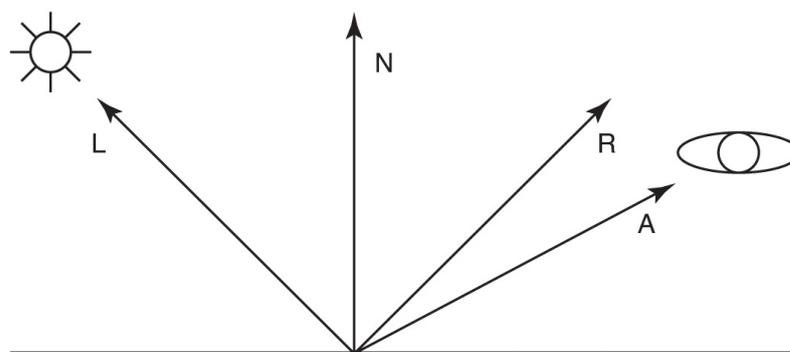
- Das Beleuchtungsmodell von Phong setzt sich damit aus diesen drei Komponenten zusammen

$$I = I_a + I_d + I_s$$

$$I = k_a I_{al} + k_d I_L \langle L, N \rangle + k_s I_L \langle R, N \rangle^n$$

- Vereinfachende Darstellung mit Grauwerten statt Farbwerten
- Einfache Erweiterung: Arbeit in drei Kanälen (R, G, B)

## Phong: Winkel



**Abbildung 8.12:** Vektoren zur Berechnung des Beleuchtungsmodells nach Phong: L zeigt zur Lichtquelle, N ist die Flächennormale, R der reflektierte Lichtstrahl und A zeigt zum Auge.

## Verdeckung: Z-Buffer

- Für jeden einzelnen Bildwert nicht nur den Farbwert RGB speichern, sondern auch die Tiefeninformation Z
- Z-Buffer initialisiert mit größtmöglicher Entfernung zur Kamera
- Beim Zeichnen eines Polygons wird die aus den Objekteigenschaften und dem Licht berechnete Farbe eingetragen, plus Z-Wert
- Steht hier schon eine kürzere Entfernung wird das Pixel nicht gezeichnet
- Etwas komplexer bei Transparenz
- Kritisch ist die numerische Genauigkeit
- Wegen effizienter Berechnung Rechnen z.B. mit Festkommadarstellung, 16 Bit
- Sind hintere und vordere Schnittebene zu weit voneinander entfernt und liegen Polygone auf gleicher Z-Ebene kann es wg. Rundungsfehlern zu Artefakten kommen

## Übersicht lokale Shader

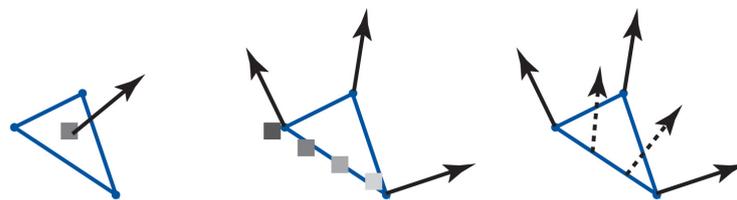


Abbildung 8.14: Funktionsweise des Flat Shading, Gouraud Shading und Phong Shading

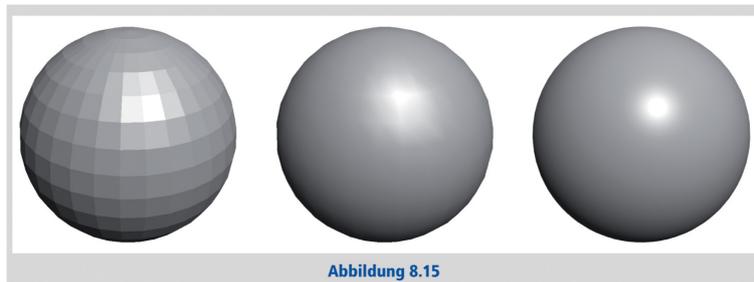


Abbildung 8.15

## Flat Shading

- Eine einzige Flächennormale für jedes Polygon
  - Beispielsweise das Kreuzprodukt zweier Kanten
- Mit dieser Flächennormalen sowie den Vektoren zur Kamera und zum Licht sowie der Oberflächenfarbe wird eine einfache Beleuchtungsfunktion berechnet
- Dies kann z.B. das eingeführte Model von Phong sein
- Die ermittelte Farbe gilt für alle Pixel des Polygons
- Pro Polygon muß die Beleuchtungsfunktion nur einmal ausgewertet werden
- Der entstehende Bildeindruck ist entsprechend kantig

## Gouraud Shading

- an allen Eckpunkten des Polygons wird eine Beleuchtungsfunktion ausgewertet, und zwar mit den jeweiligen Flächennormalen
- So ermittelte Farben werden entlang der Kanten linear interpoliert
- Über die Polygonfläche ergibt sich ein Farbverlauf
- Da benachbarte Polygone an den Eckpunkten die gleichen Normalenvektoren haben sind die Übergänge geglättet
- Für jedes Polygon muß die Berechnung aber mehrfach durchgeführt werden
- Dazu kommt die Interpolation der Farbe für jedes Pixel

## Phong Shading

- Hier wird der Normalenvektor über die gesamte Polygonfläche hinweg interpoliert
- Erheblich höherer Rechenaufwand durch die Interpolation von Vektoren
- Wesentlicher Vorteil bei der Darstellung von glänzenden Oberflächen
- Je genauer die Approximation an den Ausgaben eines Glanzlichts ist, desto eher wird ein relativ kleines Glanzlicht auch aufgenommen

## Schatten bei lokalen Shadern

- Für die bisher beschriebene Modellierung der Lichts brauchen wir für die diffuse Reflexion und die Glanzreflexion den Winkel zur Lichtquelle
- Problem: Auch die Lichtquelle kann analog zu Objekten von anderen Objekten verdeckt werden
- In diesem Fall sollte das Licht nicht modelliert werden
- Lösung: Einsatz des Z-Buffers, wobei dieses mal von der Lichtquelle aus gerechnet wird
  - Alle Objekte werden von der Lichtquelle aus gesehen in einen Z-Buffer geschrieben
  - Nur bei dem Objekt, welches der Lichtquelle am nächsten ist, wird die lokale Beleuchtung (z.B. mit Phong) berechnet

## Raytracing

- Verfolgung je eines Blickstrahls vom menschlichen Auge durch jedes Bildschirmpixel in die 3D-Szene
- Trifft auf vordersten davon geschnittene Objektfläche
- Dort entweder reflektiert, gebrochen oder absorbiert
- Bis zu einer gewissen Reflexionstiefe werden gebrochene/reflektierte Streifen ebenfalls verfolgt
- Beim Rücklauf durch die Rekursion ergibt sich die Pixelfarbe als Summe aller gebrochenen und reflektierten Strahlen sowie ggf. der Oberflächenfarbe am Schnittpunkt
- Die Farbe wird dabei durch Beleuchtungsfunktionen analog zu den lokalen Verfahren berechnet
- Nebenbei leistet Raytracing einen Beitrag zur Elimination von Objekten, die im Bild nicht auftauchen können
- Statt Z-Buffer tauchen nur Objekte auf, die von einem Lichtstrahl getroffen werden, der in der Kamera endet

## Radiosity

- In der Natur wird jede Oberfläche nicht nur direkt von einer Lichtquelle beleuchtet, sondern auch durch Lichtanteile, die von anderen Objekten diffus reflektiert wurden und jetzt indirekt auf die Oberfläche einwirken
- Oberflächen werden unterteilt
- Dann wird die direkt einfallende Lichtenergie berechnet
- Daraus und dem diffusen Reflexionskoeffizienten ergibt sich die abgestrahlte Menge
- Dann wird iterativ berechnet, welcher Anteil der Lichtenergie bei wieder anderen Oberflächen ankommt
- Nahe beieinander liegende Flächen beeinflussen sich stärker als weit entfernt liegende
- Nach vielen Durchläufen stabilisiert sich das Verfahren
- Berechnung aufwendig, quadratisch zur Anzahl der Flächen und iterativ
- Radiosity ist unabhängig von der Kameraposition!

## 11 Texte und Typographie

### Beschreibungsebenen in der Linguistik

- In einer bestimmten Sprache verfaßter Text kann auf verschiedenen Ebenen betrachtet werden
- Wörter einer Sprache sind in einem **Lexikon** zu finden
  - Lexikalische Ebene
- Korrekte Schreibung eines Wortes ist die Orthographie
  - Oder, in "reformierter neuer Rechtschreibung", Orthografie
  - Lexikalischer Fehler: Versendung eines falschen Wortes
  - Die falsche Zusammenstzung von Buchstaben ist ein orthographischer Fehler
- Regeln, nach denen korrekte Sätze (Folgen von Wörtern) gebildet werden, bilden die **Grammatik**
  - Ein grammatischer Fehler z.B. eine falsche Satzkonstruktion wäre
  - Ein lexikalischer Fehler wäre z.B. die falsche Wahl eines Wertes

### Ebenen (contd.)

- Lexikon und Grammatik beschreiben den formalen Aufbau, oder die **Syntax**
- Korrekte Syntax läßt sich auch ohne Verständnis der Bedeutung überprüfen
  - Rechtschreib- und Grammatikprüfung in Textverarbeitungen
  - Lexiko-grammatisch korrekte Sätze können sinnfrei oder mehrdeutig sein
- Die **Semantik** betrachtet den Bedeutungsinhalt eines Satzes
- Dabei werden Wörtern und Sätzen einer Sprache Dinge, Vorgänge oder Sachverhalte in der realen Welt zugeordnet
  - Für formale Sprachen kann dies vollständig und korrekt geschehen
  - Formale Semantiken sind i.d.R. rekursiv definierbar und lehnen sich an die rekursive Beschreibung der Syntax an
  - Die formale Ableitungsregel innerhalb einer Grammatik bestimmt auch die zugehörige Semantik

## Semantik natürlicher Sprachen

- Die Semantik natürlicher Sprachen ist häufig mehrdeutig und hochgradig abhängig vom Kontext
  - “Ich fange an zu arbeiten”
    - \* Student zu seinen Eltern
    - \* Arbeitsloser beim Jobcenter
    - \* Arbeitender zu seinem Chef
- Semantik natürlicher Sprachen nur in Ausschnitten formal beschreibbar
  - Erkennung und Generierung von Textfragmenten in bestimmten, abgegrenzten Bereichen
- Die **Pragmatik** beschreibt schließlich die Verwendung von Sprache zur Erzielung bestimmter Wirkungen
- Sätze eines Textes können bspw. nach ihrer Funktion als Sprechakte beschrieben werden
  - Frage
  - Behauptung
  - Bestätigung

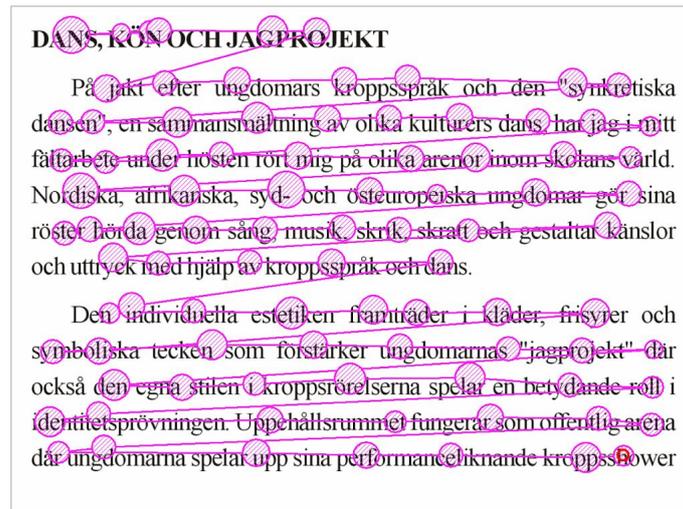
## Ebenen natürlicher Sprache

- Klassische semiotischen Ebenen
  - Zeichen
  - Syntax
    - \* Lexis
    - \* Grammatik
  - Semantik
  - Pragmatik
- Es gibt auch linguistische Theorien, die eine andere Stratifizierung vornehmen

## Wahrnehmung von Texten: Lesen

- Lesen komplexer Vorgang, der nicht so linear abläuft, wie man naiv annehmen könnte
- Auge sieht nur in einem engen Bereich von wenigen Grad wirklich scharf
  - Fovea Centralis: Punkt des schärfsten Sehens
    - \* Etwa 1.5-2° um die anatomischen Achse des Auges
  - Lesebereich: etwa 5° um die anatomischen Achse
  - Keine Stäbchen zum Helligkeitsehen, nur Zapfen
  - Foveales oder zentrales Sehen: der hier fixierte Ausschnitt
- Um einen größeren Bereich abzudecken, befindet sich das Auge in ständiger Bewegung
- Sakkaden: Kleine, schnelle Sprünge (etwa 50ms)
- Fixationen: dazwischen liegende Ruhepausen, beim Lesen etwa 250-400 ms

## Sakkaden



Wikipedia

## Verarbeitung

- Während der Fixation wird ein scharfes Bild vom Auge aufgenommen, das danach analysiert wird
- Während der Sakkaden sind wir durch die Bewegungsunschärfe eigentlich "blind"
- Die Verarbeitung des "aufgenommenen" Bildes passiert jedoch weiter
- Aus den vielen aufgenommenen Einzelbildern und dem Ergebnis des peripheren, also nicht fovealen, Sehens rekonstruiert das Gehirn den Eindruck eines rundum scharfen, überall farbigen Bildes

## Textlesen

- Beim Lesen westlicher Texte folgen die Sakkaden in etwa den Textzeilen von links nach rechts und von oben nach unten
- Am Zeilenende wird an den Anfang der nächsten Zeile zurückgesprungen
- Bei durchschnittlichen Schriftgrößen und Leseabständen sehen wir einen Bereich von etwa 5 Buchstaben scharf
- Reicht aus, um einfache und bekannte Wörter zu erkennen
- Komplizierte Wörter brauchen mehrere Fixationen
- Bei schnellem Lesen und erfahrenen Lesern wird nicht auf jedes einzelne Wort fixiert
  - Viele Wörter lassen sich aus der Grammatik und dem Sinnzusammenhang vorhersagen
  - Deren Bild kann durch peripheres Sehen überprüft werden

## Regression

- Gelegentlich springt das Auge auch zurück zu bereits gelesenen Stellen (**Regression**)
- Dies mag mit unserem strukturellen Verstehen von Sätzen oder Satzteilen zusammenhängen (cognitive process model)
- Lesen ist nach diesem Verständnis ein aktiver Wahrnehmungsprozeß
  - Die erste Phase, die Wahrnehmung durch das Auge, liefert nicht nur Informationen für spätere Phasen, sondern wird von diesen (z.B. der Worterkennung) auch gesteuert
- Anzahl der Regressionen steigt bei komplizierten Texten an
- Liefert Aussage darüber, wie schwer das Textverständnis für den einzelnen Leser ist

## Subvokalisierung

- Bei der Subvokalisierung werden beim Lesen die Laute der gelesenen Texte stumm erzeugt
  - Vermindert den Cognitive Load
  - Beinhaltet Muskelbewegungen des Kehlkopfes, die vom Lesenden i.d.R. nicht wahrnehmbar sind
  - Nach Baddeley und Hitch wesentlicher Bestandteil der “Phonological Loop”: Subvokalisierung ermöglicht es, visuell aufgenommenen Informationen im Auditory Sensory Store zu verarbeiten
- Vollständige Elimination der Subvokalisierung scheint unmöglich zu sein,
  - Allerdings: fMRI-Studien zeigen Unterschiede der aktivierten kortikalen Regionen zwischen schnellen und langsamen Lesern, insbesondere in Regionen zur Spracherzeugung

## Mikrotypografie: Schriftarten und Buchstaben

- Die **Mikrotypografie** befaßt sich mit der kleinräumigen Gestaltung des Textsatzes
  - Form einzelner Buchstaben
  - Deren Anordnung zueinander
  - Kombination zu neuen Formen
- Bei der digitalen Repräsentation werden diese in einer Schriftartdatei festgelegt
- Der gleiche Text kann in verschiedenen Schriftarten gesetzt werden
- Die gleiche Schrift kann in verschiedenen Schriftartdateien beschrieben werden
  - Verschiedene Digitalisierungen der gleichen Schrift, z.B. Zapf Optima und URW Classico
  - Verschiedene Kodierungen der graphischen Form, z.B. Bitmaps, TrueType, PostScript und OpenType

## Begriffe



Abbildung 5.1: Serifen sind Verzerrungen an den Enden der Linien eines Buchstaben und erhöhen die Lesbarkeit und Lesegeschwindigkeit. Eine Ligatur fasst zwei Buchstaben für eine bessere Darstellung zu einem neuen zusammen.

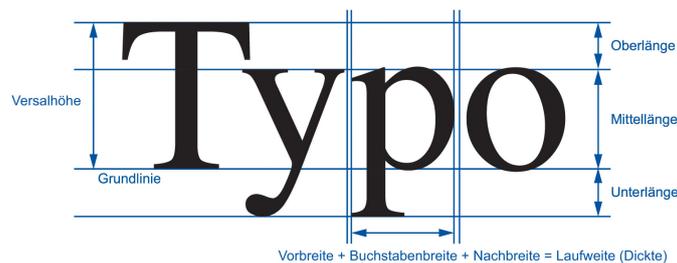


Abbildung 5.2: Einige grundlegende Begriffe in der Mikrotypografie

## Schriftschnitte

- Eine Schriftart kann in unterschiedlichen Schnitten vorliegen
- Im allgemeinen finden sich normale (regular), fette (bold) und entweder kursive (italic) oder schräge (slanted), auch in Kombination (fette kursive)
- Hochwertige Schriften können in vielen weiteren Schnitten vorliegen (light, black)

Helvetica Courier Times  
**Helvetica Courier Times**  
*Helvetica Courier Times*  
***Helvetica Courier Times***

Abbildung 5.3: Die Schriftarten Helvetica, Courier und Times, jeweils in den Schnitten normal, fett, kursiv und fett kursiv

## Kerning

- Die Harmonie eines gesetzten Textes kann verbessert werden, indem Buchstaben je nach ihrer Form näher zusammengedrückt oder weiter auseinander gezogen werden
- Dieses harmonische Anordnen heißt auch Kerning

Tat The

Abbildung 5.4: Im Kerning werden Buchstaben je nach ihrer Passform näher zueinandergerückt oder weiter auseinandergezogen

## Durchschuß und Zeilenabstand

Im Bleisatz wird die  
 Gliederung der Zeilen  
 über den *Durchschuß* geregelt.  
 Im Gegensatz dazu wird  
 beim heute eingesetzten  
 Fotosatz der *Zeilenabstand*  
 verwendet.

Quelle: Birgit Lugin

## Zeilenlänge

- Bei der Länge der einzelnen Zeilen ist Rücksicht auf das menschliche Sehen zu nehmen
- Wir haben gesehen, daß es einzelne Fixationspunkte gibt, bei denen der Mensch eine Stelle im Text genauer anschaut
- Wenn wir zu viele Fixationen pro Zeile haben, kann der Rücksprung zum Zeilenanfang schwierig werden
- Beste Lesbarkeit bei etwa 50-70 Zeichen pro Zeile
- D.h. daß die Zeilenlänge eine Funktion der Schriftart und der Schriftgröße ist
  - Festlegung von Seitenrändern ohne passende Festlegung der Schriftart und -größe typografisch zweifelhaft
  - Schriftgrößen um die 10-11 Punkt auf A4 mit Rändern von 1.5-2cm führen zu überlangen Zeilen

## Makrotypografie: Gestalten mit Schrift

- Makrotypografie: Anordnung der Textzeilen auf einer Druckseite oder dem Bildschirm
- Äußere Begrenzung des mit Bildern oder Text bedruckten Bereichs heißt Satzspiegel
- Die Teile zwischen dem Satzspiegel und dem Rand sind die Margen
- Maßgeblich für harmonische Erscheinung verantwortlich
- Verschiedene Verfahren für die Bestimmung des Satzspiegels
  - Goldener Schnitt
  - Satzspiegel nach Tschichold

## Goldener Schnitt

**Goldener Schnitt:** Als goldenen Schnitt bezeichnet man die Teilung einer Strecke derart, daß sich der kleinere zum größeren Teil so verhält wie der größere zur Gesamtlänge, also  $a : b = B : (a + b)$ . Dieses Verhältnis beträgt etwa 1:1,618 und wird in unserem Kulturkreis seit der Antike als besonders harmonisch empfunden. Der goldene Schnitt wird z.B. in der Architektur, Skulptur, Malerei.

## Satzspiegel: Konstruktion

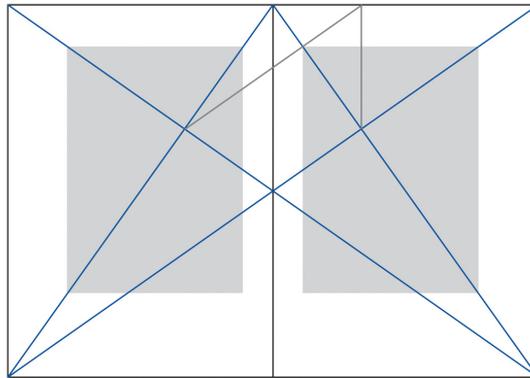


Abbildung 5.5: Konstruktion eines harmonischen Satzspiegels für eine gedruckte Doppelseite, z.B. in einem Buch

## Breite des Satzspiegels

- Aufgrund der wahrnehmungspsychologischen Grundlagen ergibt sich, daß die Breite des Satzspiegels (also die Zeilenlänge) nicht willkürlich gewählt werden sollte
- Eine Möglichkeit der Berechnung:
  - Bestimme Schriftart und Schriftgrad
  - Berechne die Breite eines Textes von 50-70 Zeichen
  - Setze den Satzspiegel dementsprechend
  - Wähle die Position des Satzspiegels
  - Dadurch bestimmt sich die Anzahl der pro Seite möglichen Zeilen
- Grundsätzliche Vorgehensweise bei den KOMA-Script  $\LaTeX$ -Paketen

## ASCII

- Die auftretenden Buchstaben, Ziffern und Zeichen (Hexadezimal 20 bis 7F) entsprechen weitgehend den Tasten einer amerikanischen Tastatur
- Steuerzeichen 00 bis 1F ursprünglich zur Fernsteuerung von Druckern und Fernschreibern (LF, CR) oder Bandgeräten (EM, FS)
  - Auf Textkonsolen übertragen, wo sinnvoll
- Da in ASCII keine internationalen Zeichen (deutsche Umlaute) darstellbar sind wurden abwärtskompatible Erweiterungen definiert
  - ISO 8859-1 wurde 1986 eingeführt und enthält Umlaute und andere länderspezifische Zeichen Westeuropäischer Sprachen
  - Verwendet 8 Bit und bettet die druckbaren ASCII-Zeichen in den unteren 7 Bit ein
  - Typographisch interessante Zeichen z.B. das nicht umbrechbare Leerzeichen (NBSP) oder der bedingte Trennstrich (SHY)

## Unicode

- Die so verfügbaren Zeichen stoßen recht schnell an prinzipielle Grenzen
- Austausch mit anderen, inkompatiblen Erweiterungen, die den gleichen Adreßraum belegen, erschwert
- Besonders schwierig wenn ganze Alphabete (kyrillisch, griechisch) dargestellt werden sollen
- Erste Version des Unicodes (ISO 10646) 1991
  - Umfaßte europäische, nahöstliche und indische Schriften
- Ziel des Unicodes ist es, alle auf der Welt verwendeten Schriftzeichen in einer einzigen Codierung abbilden zu können
  - Auch ikonische Systeme, wie Emotes
- Ständig weiterentwickelt und abwärtskompatibel um neue Zeichen ergänzt
- Verwendet derzeit 17 Ebenen zu 16 Bit, etwa 1 Million möglicher Werte
- Hexadezimalwerte mit führendem U+, U+00DF ist ß

## UTF-8

- IETF: soll Standard für alle Internet-Protokolle sein
- UTF-8 benutzt eine variable Wortlänge von 8 bis 32 Bit
- Bei Verwendung von ASCII-Zeichen kein erhöhter Speicherbedarf (8 Bit Worte)
  - U+0000 - U+007F (ASCII) wie ASCII
  - U+0080 - U+07FF Folge von 2 Bytes, 110xxxxx 10xxxxxx
  - U+0800 - U+FFFF Folge von 3 Bytes, 1110xxxx 10xxxxxx 10xxxxxx
  - U+010000 - U+1FFFFF Folge von 4 Bytes, 11110xxx 10xxxxxx 10xxxxxx 10xxxxxx
- Übergeordnetes Prinzip: Zeichen, die mehr als 7 Bits zur Codierung benötigen, werden mit einem Startbyte und mehreren Folgebytes codiert
- Startbyte beginnt entweder mit einer 0 und ist dann das einzige Byte, oder mit einer 1 und enthält danach so viele 1, wie Folgebytes kommen, abgeschlossen mit einer 0
- Folgebytes beginnen immer mit einer 10

## Bitmap-Fonts

- Darstellung der einzelnen Buchstaben als Rastergraphiken (Bitmaps)
- Diese wurden einfach an die entsprechende Speicherstelle kopiert
- Sehr schnelle Darstellung ohne großen Rechenaufwand
- Wesentlicher Nachteil: Für jede Größe und jede Auflösung des Zielgerätes war ein gesonderter Bitmap-Font erforderlich
- Nicht ästhetisch zufriedenstellend skalierbar
- Auch heute noch als Konsolenschrift antreffbar
- Diese Methode wurde vor allem mit dem Auftauchen des Desktop Publishing (DTP) unbrauchbar
- Übergang, Schriftarten durch Linien- und Kurvenzüge zu beschreiben
- Solche Schriften werden auch Vektor-Fonts genannt

## Vektor-Fonts

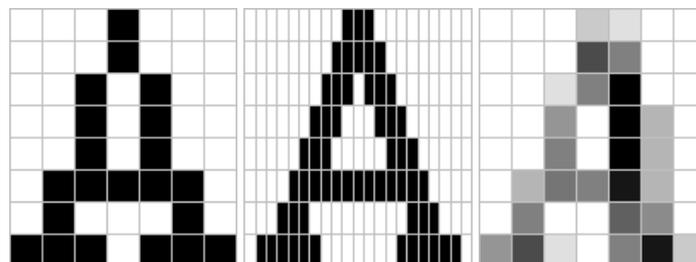
- Im wesentlichen vier Standards
  - TrueType
  - PostScript
  - OpenType
  - Web Open Font Format
- Alle beschreiben die einzelnen Zeichen mittels Interpolationskurven und enthalten Kerning-Tabellen
  - TrueType: quadratische Bézier-Kurven
  - PostScript: kubische Bézier-Kurven, durch die höhere Anzahl von Kontrollpunkten können diese Schriften die ursprüngliche (Bleisatz) Form besser annähern
  - OpenType können wiederum TrueType oder eine Variante von PostScript Beschreibungen enthalten
  - WOFF ist im wesentlichen mit Metadaten annotierte und komprimiertes Open- oder TrueType

## Aliasing

Medieninformatik  
Medieninformatik

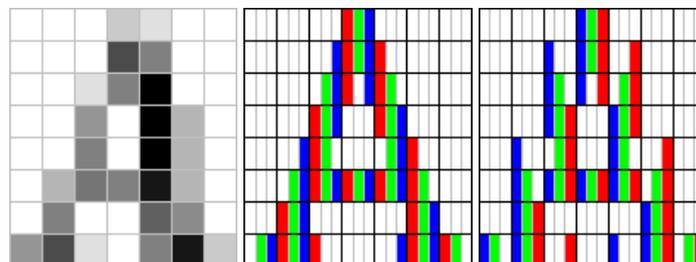
Abbildung 5.11: Font-Rendering ohne (oben) und mit Antialiasing (unten)

## Subpixel-Rendering



Links: SW-Rasterisierung mit gegebener Pixelzahl  
Mitte: SW-Rasterisierung mit dreifacher horizontaler Pixelzahl  
Rechts: Antialiasing durch Graustufen

## Subpixel-Rendering (contd.)



Links: Antialiasing durch Graustufen  
Mitte: Antialiasing durch Ausnutzen von RGB-Pixeln  
Rechts: Subpixel-Rendering mit falscher Pixelreihenfolge

## 12 Multimodalität

### Medialität, Codalität und Modalität

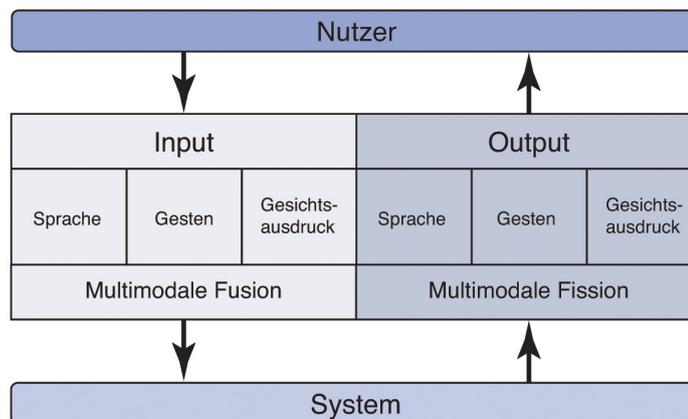
**Multimedialität:** Systeme, in denen mehrere Medientypen (wie Text, Bilder, Videos) benutzt werden, bezeichnet man als *multimediale* Systeme

**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

### Multimodale Systeme



**Abbildung 9.2:** Prinzip symmetrischer multimodaler Interaktion bei SmartKom. In diesem System wurden die gleichen Modalitäten für Ein- und Ausgabe genutzt.

### 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

## 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)

## 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 vermehrt 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

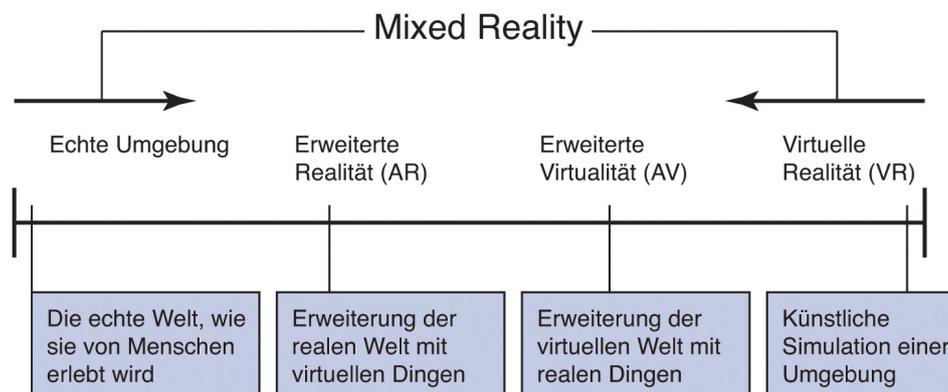
- Die skizzierte Verarbeitungskette aus
  - Phonemerkennung
  - Worterkennung
  - Satzerkennung
  - Bedeutungserkennungist stark verkürzt
- In der Praxis notwendige Teilschritte können umfassen

- *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)

## 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.

## Virtual Reality: 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. andere Konfigurationen möglich (Sphäre). 3D-Darstellung durch Shutterbrillen oder Polarisation.

## 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

## 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

## 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*)

## 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.

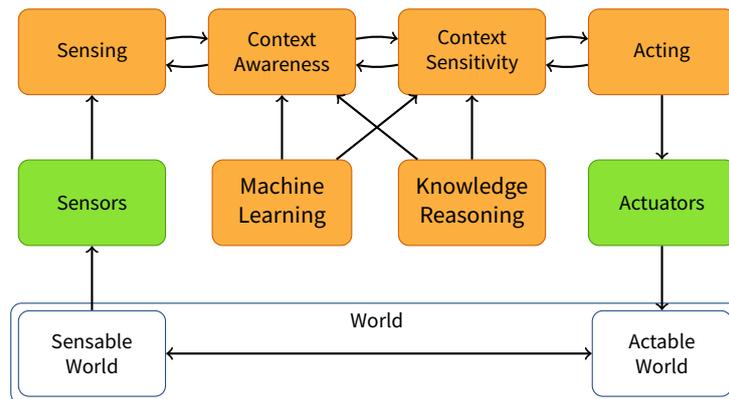
- **Example:** the same system fetches the newest versions of electronic patient records of all patients in the room from the hospital systems. When the user stands close to the bed of a patient, the system automatically displays them.

## 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

## Architectures: Ambient Intelligence



## Authoring Digitaler Medien

- 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)

## 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

## 13 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.

### References

### Literatur

Dale Purves, George J Augustine, David Fitzpatrick, Lawrence C Katz, Anthony-Samuel LaMantia, James O McNamara, and S Mark Williams., editors. *Neuroscience, 2nd Edition*. Sinauer Associates, Sunderland (MA), 2001.

John Watkinson. *The MPEG Handbook: MPEG-1, MPEG-2, MPEG-4*. Technology & Engineering. Taylor & Francis, 2004.

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>.