

1 Text

1.1 Schrift und Text als Medium

Schrift ist eine visuelle Darstellungsform der Sprache, zusammengesetzt aus *Schriftzeichen*. *Text* ist „Gewebe der Sprache“.

Wird eingegeben/weitergegeben über Tastatur, Handschrift, Sprache ... und ausgegeben über Monitor, Drucker, ...

Speicherung auf physischen Medien wie Festplatten, Papier, Stein, Sand, ...

Syntax und *Semantik* beschreiben Struktur, Inhalt und Bedeutung von Text.

1.2 Rechnerinterne Darstellung von Text

Kenntnis über Aufbau und Einsatz unterschiedlicher Zeichensätze (ASCII, UCS, Unicode, ...)

Das Kodierungsmodell definiert einen konzeptionellen Rahmen zur Abbildung der Menge der abstrakten Zeichen auf eine konkrete Kodierung als Bitfolge und dessen interne Speicherung.

Zeichensatz ist eine rechnerinterne Darstellung einer Sammlung von Zeichen. Notwendigerweise muss die Anzahl der verfügbaren Bits hinreichend groß sein um alle verwendeten Zeichen eindeutig abbilden zu können.

Wünschenswerte Attribute der *Abbildungsfunktion*:

1. Codierung in kleinstmögliche Zahlen/Bitfolgen
2. Wohldefinierte Ordnung, etwa anhand des Alphabets
3. Standardisierung, weltweite Einigung auf eine Abbildungsfunktion

Die Abbildungsfunktion ist gespeichert als Form einer Codetabelle (code table, character set, ...) bei dem jedes Zeichen eine Nummer (code point, position, ...) erhält.

ASCII ist ein 7-Bit Zeichensatz ($2^7 = 128$), das achte Bit ist immer null.

Erweiterungen in Form von länderspezifischen Varianten (ISO DE/US) und auf 8 Bit (DOS, ANSI).

Unicode Abbildung aller Schriftzeichen der Welt in einem Zeichensatz (system-, programm- und sprachunabhängig, klarerweise).

Anfangs 16-Bit, mittlerweile 21-Bit (?) und >140k belegte Zeichen. Die Unicode Notation hat die Form $U + 0000$ wobei 0000 durch den Codepoint zu ersetzen ist. Wrapper über ANSI und ISO.

Kodierungsformat Angabe, wie das Bitmuster der Codetabelle tatsächlich intern gespeichert wird.

Probleme der naheliegenden kanonischen Speicherung (direkt, je 8 Bit in einen Byte speichern) sind Redundanz durch lange Nullfolgen, ebenso ist ASCII nur 7-Bit.

Zeichensätze wie *UTF* definieren ein Kodierungsformat. UTF unterstützt etwa variabel lange Codewörter. (UTF-8: 1-4 Bytes, UTF-16: 2 oder 4, UTF-32 fix 4 Bytes, kanonisch.)

Fonts Die *Schriftart* (Typeface) bestimmt das Erscheinungsbild des darzustellenden Zeichens. Der *Schriftsatz* bestimmt die rechnerinterne Darstellung der Erscheinungsbilder, etwa durch Bitmaps (Bild, bestehend aus einzelnen Punkten) oder Vektoren (ähnlich SVG).

Bitmap fonts ist an eine bestimmte Auflösung gekoppelt/angepasst, Skalierung auf unvorhergesehene Auflösungen führt zu Aliasing („Verpixelung“) geht aber schnell.

Outline Fonts sind wie alle Vektorgraphiken unabhängig von Auflösungen aber aufwändiger in der Handhabung (Kerningtable, rendering Hints, ...).

2 Bild

2.1 Wahrnehmungspsychologische Grundlagen

Licht ist Teil des elektromagnetischen Spektrums, Menschen haben einen Tri-Chromatisches Sehsinn (Blau – kürzere Wellenlängen, Grün – mittlere, Rot – längere). Etwa 6 Million solcher farbempfindlichen Zapfen.

2.2 Farbmodelle

RGB Prinzip der *additiven* Farbmischung durch Angabe der Helligkeit des Rot-, Grün- und Blauanteils. Jeder Anteil wird einzeln quantisiert. 24 bit Farbtiefe führen zu 8 bit je Grundfarbe (also 0-255 pro Farbanteil). Geräteabhängiges Farbmodell: der Wert ist zwar exakt definiert, tatsächliche Darstellung ist allerdings Geräteabhängig.

CMY Prinzip der *subtraktiven Farbmischung*. Mischen von Körperfarben basierend auf der Absorption von Licht, bestimmte Wellenlängen werden herausgefiltert, nicht absorbierte Farben werden reflektiert. Cyan absorbiert Rot, Magenta Grün, Gelb Blau. Abhängig von bestrahlendem Licht.

YUV Aufgebaut aus *Luminanzsignal* Y und zwei *Chrominanzsignalen* I/U und Q/V. Verwendung in der Fernsehtechnik zur Steigerung der Übertragungseffizienz. Nur Luminanzsignal: Schwarz-weißfernsehen. Luminanz ist wichtiger für Detail-Eindruck als Farbinformation und erhält deswegen mehr Bandbreite.

Colorsampling ist sinnvoll um Speicher zu sparen. Unterschiedliche Zahl von Bits pro Farbkanal. 4:4:4 keine Reduktion, 4:2:2 2/3 Reduktion, 4:2:0, 4:1:1.

Color Lookup Table ist sinnvoll zum Speichersparen insbesondere beim RGB-Modell. Pixel enthalten Index von Tabelleneintrag in dem die jeweilige Farbe gespeichert ist. GIF etwa 256 Farbeinträge.

Bitmaps/Rastergraphiken sind einfache Matrix deren Einträge der Pixelwert an der jeweiligen Stelle ist (üblicherweise RGB).

2.3 JPEG

1. Zerlegung in Farbkanäle (YCbCr-Farbraum), einfache Matrixmultiplikation von vordefinierten Werten mit den RGB Werten.
2. Downsampling (Chroma-Subsampling) der Farbinformationen (4:2:2, in Spalten je zwei Y ein U/V oder 4:2:0, zentral je 4 Y ein U/V).
3. Je Kanal Zerlegung in 8x8 Blöcke
4. Für jeden Block:
 - (a) Diskrete Kosinustransformation. Umwandlung in DCT-Koeffizienten, größere Werte sind ‚oben links‘.
 - (b) Quantisierung. Die 8x8 DCT-Koeffizienten werden mit Normalisierungstabelle verknüpft (extern vorhanden bzw. muss mit Bild mitgeliefert werden).
 - (c) Entropiecodierung. Standardmäßig Huffman-Kodierung, verlustfrei, Präfix-Code, Quellsymbolen werden Codewörter mit verschiedener Länge zugeordnet. Häufiger auftauchende Zeichen werden mit weniger bits repräsentiert als seltener auftauchende.

Zick-Zack Scan DCT-Koeffizienten werden durch Differenzbildung zu den Koeffizienten des vorhergehenden Blocks codiert und durch eine Zickzack Sequenz angeordnet (0,0), (1,0), (0, 1), (1, 1), (2, 1),... Lange Nullketten sind zu erwarten.

JPEG 2000 Bessere Bildqualität, kürzere Kodier- und Dekodierzeit, niedriger Speicherbedarf. Transformation (Wavelets anstelle von DCT), angewendet auf ganzes Bild (nicht auf Block).

3 Audio

3.1 Wahrnehmungspsychologische Grundlagen

Schall sind durch Energie hervorgerufen *Schwingungs-* und Wellenvorgänge, verbreiten sich in einem elastischen Medium. (Kinetische Energie beim Anschlag einer Klaviertaste erzeugt Vibration der angeschlagenen Saite, Vibration wird durch den Resonanzraum des Klaviers verstärkt und breitet sich in der Luft aus.)

Dargestellt als Sinusschwingung die Ton einer bestimmten Frequenz entspricht, Wellenlänge ist abhängig vom Medium.

Die *Tonhöhe* ist bestimmt durch die Frequenz (Schwingungen pro Sekunde). Der Hörbereich des menschlichen Ohrs liegt zwischen 16 und 20000 Herz. Je höher der Ton, desto höher die Frequenz.

Der Schalldruck ist ein objektives physikalisches Maß (in Pascal, [pa]) oder normiert auf hörbaren Bereich [0 ... 130dB].

Primärempfindungen sind

- Tonhöhe (versch. Klaviertasten)
- Lautstärke (Boeing vs. Cessna)
- Klangfarbe (etwa gleicher Ton auf versch. Instrumenten)

Frequenzmaskierung Laute Töne bestimmte Frequenz machen leisere Töne ähnlicher Frequenz unhörbar.

Zeitliche Maskierung Laute Töne beeinflussen Wahrnehmung frequenzähnliche Töne in direkter zeitlicher Nachbarschaft.

4 Rechnerinterne Darstellung

1. Aufzeichnung von Audiosignalen. Mikrophon wandelt Schallwelle in elektrisches Signal, Spannungsverlauf ist analog zum Druckschwankungsverlauf an Membran des Mikrofons.

An.: Analoges Signal speichern auf Magnetband, Schallplatte, ...

Dig.: Analoges Signal mittels A/D Konverter in binäre Darstellung umwandeln und speichern. Mittels D/A Konverter wieder als Analogsignal an Lautsprecher weitergeben.

Sollte folgende Eigenschaften berücksichtigen

- Fidelity. Wahrgenommene Qualität vs. Originalsignal. (Abhängig von Abtastrate und Quantisierungsstufen, mehr ist besser.)
- Bitrate. Bits pro Zeiteinheit des Codierungsergebnisses. (Abhängig von A. und Q., weniger ist besser.)
- Komplexität. Benötigter Aufwand (HW/SW) zur En- und Dekodierung, oft mit Echtzeitanforderung.

Pulse Code Modulation Abtasten und Quantisieren der Amplitude des Originalwerts. Hoher Bandbreitenbedarf:

$$\frac{\text{bitrate}}{\text{s}} = \frac{\text{Abtastrate}}{\text{Hz}} \cdot \frac{\text{Bits}}{\text{Abtastwert}} \cdot \text{Kanäle}$$

Wichtig ist Wahl der Abtastrate (Nyquist Theorem)

$$\text{Abtastrate} > 2 \cdot \text{max. Frequenz}$$

Bei der Quantisierung werden abgetasteten Spannungswerten diskrete Zahlenwerte zugeordnet, etwa 8 oder 16 bit.

4.1 Kompression

- Verlustbehaftete Veränderung der Sampling/Quantisierungsparameter. (Weniger Kanäle, weniger Bits pro Sample, weniger Samples/kleiner Samplingfrequenz.)
- Quellcodierungen (Differential PCM, Deltamodulation, Dynamikkompression)

Bei der Differential Pulse Code Modulation wird nur die Differenz zwischen jeweils aufeinanderfolgenden Codeworten gespeichert. Üblicherweise gibt es geringere Differenzen zwischen Nachbarn bzw. allgemein kleine Differenzen auf relativ hohem Wertebereich. Pegelverlust (weniger Quantisierungsbits) schlechtere Übertragung von Höhen. (Gibt adaptive Variante mit anpassbarer Anzahl von Quantisierungsbits, etwa zwei Modi.)

Bei Delta-Modulation werden Differenzen mit einem Bit kodiert, entweder Erhöhung oder Senkung um gegebenes (sehr kleines) Delta. Erfordert hochfrequente Abtastung um „mitzukommen“.

Dynamikkompression (nichtlineare Quantisierung) baut auf dem Weber-Fechnerschen Grundgesetz der Psychophysik auf: „Ein Reiz muss gegenüber einem Schwellenreiz logarithmisch wachsen, wenn er als stärker empfunden werden soll.“ Leise Geräusche werden also differenzierter wahrgenommen als Laute, müssen genauer quantisiert werden. (Logarithmische Verteilung der Quantisierungsstufen.)

Psychoakustik Was nicht gehört wird, soll auch nicht kodiert werden:

- Herausfiltern von Frequenzen unterhalb der Hörgrenze.
- Filtern auf Basis von Frequenzmaskierung und zeitlicher Maskierung.
- Sub-Band Kodierung.
 1. Aufteilen des Signals in Teil-Frequenzbänder. (Ähnliche Frequenzen sind „benachbart“.)
 2. Ermittlung des Signalpegels für jedes Teilband.
 3. Für jedes Teilband wird Maskierungsschwelle berechnet, Teilbänder die von anderen überdeckt sind, werden nicht kodiert.
 4. Kodierung der Teilbänder mit jeweils angemessener Bitanzahl.

Ein *Codec* ist ein spezifischer Ansatz (Technologie, Algorithmen) zur Kodierung von (Audio-) Daten. *Dateiformate* definieren Aufbau und Struktur der gespeicherten Daten.

Codec	Format
PCM	.wav
MPEG-1 Audio Layer 3	.mp3
Advanced Audio Coding	.m4a
Vorbis	.oga
Windows Media Audio	.wma

Als *Containerformat* bezeichnet man Formate welche Aufbau und Struktur zur Speicherung verschiedener Inhalte (Medientypen) ermöglichen und etwa verschiedene Audio- und Videocodecs enthalten (MP4, OGG, ASF).

Beispiele für Verlustfreie Audiocodecs sind MPEG-4, FLAC und ALAC.

5 Video

5.1 Charakteristika von (Digitalem) Video

Zeitliche Auflösung Wahrnehmung von Bewegung. Hier werden

- physiologische (Sehzellen, Objektverfolgung, Beschleunigungswahrnehmung, Grenze 60 FPS)
- psychologische (Elimination gleichförmiger Bewegungen, Übelkeit bei Unstimmigkeit Beschleunigungswahrnehmung/optische Bewegungswechsel, Grenze 30 FPS)

Faktoren relevant.

Räumliche Auflösung Bildformat (Zeilenanzahl, Anzahl Pixel pro Zeile, Ansichtsverhältnis). Interlaced (Halbbilder werden übertragen) oder progressive (Vollbilder) Scanning.

5.2 Rechnerinterne Darstellung von Video-Daten

Digitales Video als Folge von Einzelbildern, Reihenfolge und Zeit zwischen zwei Frames muss bei Wiedergabe berücksichtigt werden.

- 576i mit PAL (720x576, 4:2:0, 25 FPS)
- NTSC 480i mit NTSC (720x480, 4:1:1, 30 FPS)
- 1080i60 HDTV (1920x1080, 4:1:1, 60/2 FPS)

Es gibt mehrere Ansätze zur Kompression.

Intra Frame Kodierung (Motion JPEG) Jedes Bild wird einzeln nach JPEG Algorithmus komprimiert, die Bildfolge wird als Video betrachtet. Keine Abhängigkeiten zwischen Bildern berücksichtigt, beliebige Positionierung innerhalb des Videos möglich. Niedrige Kompressionsrate, keine Berücksichtigung von Audio.

Inter Frame Kodierung Video wird in Bildgruppen zerlegt, erstes Bild jeder Gruppe wird unabhängig kodiert (Referenzframe, I-Frame), weitere Bilder der Gruppe werden als Differenzbilder zum ersten kodiert.

Durch Bewegungsabschätzung werden Bewegungsvektoren zwischen den Frames ermittelt. Der „Vorhersagefehler“ und diese Vektoren können nun Kodieren von Frames in einer Bildgruppe verwendet werden.

Vergleichskriterien ist etwa die mittlere absolute Differenz der Pixel, die Summe der absoluten Differenzen, quadratischer Fehler, ...

Suchstrategien sind etwa sequentielle Suche, hierarchische Suche (Start mit niedriger Auflösung, Verfeinerung der Suche in Bereich mit bestem Match), 2C-logarithmische Suche (9 Suchblöcke um den aktuellen Block herum, beim besten Match wieder von vorne mit halber Distanz)

P-Frame aus Vergleich mit vorhergehenden Bildern, B-Frame (bidirectional) aus Vergleich mit vorhergehenden und nachfolgenden Bildern.

5.3 H.261

Einsatzgebiet ist Videotelefonie, Echtzeitanforderung, bewegungsarm, kodiert Intra- und Inter-Frames. Nach Intra-Frame kommen einige Inter-Frames. Weniger Inter-Frames als üblich notwendig wäre aufgrund des Einsatzgebiets.

Grundprinzip der Digitalisierung Ansätze der Kompression: Intra- Frame-Codierung, Interframe Codierung Standards Motion JPEG, MPEG-1, MPEG-2, MPEG-4, MPEG-7, MPEG-21, H.261, H.263, H.264