

Datenreduktion (= verlustbehaftete Codierung)

- im Englischen LOSSY CODING → es gehen Informationen verloren
- zur Erinnerung: bei der Entropiecodierung (z.B. Huffman, LZW) wird die Redundanz verringert, die Information bleibt erhalten

Psychoakustisch-motivierte Datenreduktion

- alles unterhalb der Hörschwelle wird entfernt, weil unhörbar
- Trennung zwischen relevanter (wahrnehmbarer) Information und irrelevanter (nicht wahrnehmbarer) Information
 - Welche Anteile des komplexen Signals sind hörbar?
 - Vorhersage der Mithörschwellen!!
- es werden verschiedene Mithörschwellen für bestimmte Masker abgelegt
 - MP3 Player rechnet mit einer (linearen) Approximation der Mithörschwelle
- Basis bei Audiosignalcodierung → Verdeckung im Frequenzbereich

Grober Ablauf der Algorithmen:

- gleiches Prinzip → Annäherung der Mithörschwelle → Bildung einer oberen Mithörschwelle
- moderne Codecs untersuchen, was als Masker wirkt → weniger Unterscheidung zum nicht codierten Signal möglich

Analoge Datenreduktion:

- Vocoder als System zur Datenreduktion bei Sprache
- Filterband wird auf das Signal gelegt → nur noch Information über Frequenzband
- Extra Kanal für Rauschen
- Rekonstruktion per Synthesizer → generiert neues Signal
 - Nachteil: Sehr stark verfremdete Stimme, Subtext verschwindet

Digitale Datenreduktion:

- PCM → Signal wird codiert übertragen
- MP3 → Analyse/Synthese -Verfahren, sehr fein aufgelöst im Frequenzbereich
- Zerlegung in Frequenzbänder → Reduzierung der Wortbreite (variable Wortbreite bei datenreduziertem Format)

Berechnungsbeispiel CD:

16Bit @ 44,1kHz, stereo

$$R = M \cdot f_A \cdot n = 1,41 \cdot 10^6 \frac{\text{Bit}}{\text{s}}$$

Vergleich MP3:

$$128 \dots 256 \frac{\text{kBit}}{\text{s}} = 1 \dots 3 \cdot 10^5 \frac{\text{Bit}}{\text{s}}$$

- Bsp. 12dB SNR → man könnte bis 2Bit mit der Wortbreite runtergehen
 - Reduzierung der Wortbreite geht nur im digitalen Bereich
- Datenrate bei MP3 legt die Genauigkeit der Codierung fest
- Dynamik wird durch den Zwischenraum von Mithörschwelle und Schmerzgrenze ermittelt

Schema eines Codiersystems:

- Audio, PCM wird in das System geführt
- zeitliche Zerlegung in Blöcke (2...50ms)
- Aufteilung in zwei parallel ablaufende Prozesse:
 1. Zeit-Frequenz Analyse & Zerlegung → es entsteht ein zerlegtes Audiosignal
 2. psychoakustisches Modell ermittelt Mithörschwellen → Wortbreitenbestimmung
- Wortbreitenbestimmung wird auf das zerlegte Signal angewendet → Requantisierung und redundanzarme Codierung (z.B. DPCM) → es entsteht codiertes Audio
- aus der Wortbreitenbestimmung ergeben sich die Metadaten der Datei (Subcode)
- codiertes Audio und Subcode werden durch Multiplexen und eine Entropiecodierung in der Redundanz reduziert

Erklärungen zum Codiersystem:

- normalerweise gäbe es eine obere Frequenz, ab der man nichts mehr übertragen müsste
- Filterbank (30 Kanäle) entspricht einer Transformation mit 1024 Kanälen
- psychoakustisches Modell kann sich anhand der Blöcke mehrmals pro Sekunde anpassen
- MP3 ist von hinten nach vorne definiert → Decoder sind standardisiert → es gibt verschiedene Encodersysteme

Leitungscodierung

Qualitätsmerkmale von Leitungscodes:

- DC Gehalt (Offset)
- Taktgehalt → Clockgehalt
- Phasentoleranz

Einige Leitungscodes:

- NRZ Code unipolar (Non return to zero):
 - o bei einer Eins wird das positive Spannungsniveau angenommen, bei einer Null liegt keine Spannung an
 - o DC Gehalt vorhanden, sollte nur bei Lichtübertragung genutzt werden→ nicht nutzen, keine guten Codeeigenschaften
- NRZ Code bipolar:
 - o bei einer Eins wird das positive Spannungsniveau angenommen, bei einer Null das zugehörige negative Spannungsniveau
 - o mit geeigneter Kanalcodierung (gleich viele Nullen und Einsen) ist ein Offset auszugleichen, Taktgehalterhöhung auch mit geeigneter Kanalcodierung möglich (immer Nullen und Einsen im gleichen Abstand)
- NRZI (NRZ inverted):
 - o bei einer Eins findet in der Mitte der Bitzelle ein Wechsel vom vorliegenden Spannungsniveau auf dessen invertiertes Niveau statt
 - o immun gegen Phasendrehung
- RZ:
 - o Bei einer Eins wird die erste Hälfte der Bitzelle das positive Spannungsniveau angenommen, danach sinkt die Spannung auf 0. Bei einer Null wird die erste Hälfte der Bitzelle das negative Spannungsniveau angenommen, danach steigt die Spannung auf 0.
 - o maximaler Clockgehalt, Offset, keine Phasentoleranz

- BI-Phase Mark:
 - o Bei einer Eins wird nach dem ersten UI (Unit Interval, kleinste Intervalllänge des Signals) zum jeweils invertierten Spannungsniveau geschaltet, bei einer Null wird jeweils nach einer Bitzelle zwischen den Spannungsniveaus umgeschaltet
 - o phasenimmun, hoher Taktgehalt

Baudrate und Bandbreite:

- 1 Bit pro UI → FULLBAUDED
- 0,5 Bit pro UI → HALFBAUDED
 - höhere Effizienz bei Fullbauded
- Bandbreite bei der digitalen Übertragung (M, f_A, n)
 - a) Bitrate: $R = M \cdot f_A \cdot n$
 - Bandbreite wird mit höchster Frequenz verbunden (untere Frequenz nicht bei allen Codes gleich, z.B. NRZ = 0Hz)
 - BI-Phase Mark (niedrige Bandbreite) vs. NRZ (hohe Bandbreite)
 - b) max. Frequenz („Bandbreite“):
 - o bei NRZI: $\frac{R}{2}$
 - o bei BI-Phase Mark: R
 - bei zu vielen Übertragungen über eine Leitung muss man den Code evtl. wechseln (BI-Phase Mark → NRZI)
- Bandbreite von Rechtecksignalen → unendlich
- niedrigste Frequenz bei BI-Phase Mark → Hälfte der höchsten Frequenz
- durch Digitalisierung kommt man in den Megahertzbereich → Kabel beeinflussen das Signal

Möglichkeiten der digitalen Übertragung:

1. synchrone Übertragung:
 - ein Sample wird aus Gerät A augenblicklich gesendet und von Gerät B augenblicklich empfangen
 - Typische Fehler wegen nicht ganz exakt gleichem Takt in den Geräten:
 - o Drop Outs (fehlende Samples) → scharfe Knackser
 - o Glitches (verlorengegangene Samples)
 - o Drop Outs sind leichter, aber nicht immer zu hören
 - sauberere Verfahren:
 - a. externer Taktgenerator sendet Taktsignal an alle Geräte → PLL (Phase Locked Loop) in jedem Gerät wird aktiviert (WCLK = Wordclock)
 - b. Timecode Master in einem Gerät, die anderen sind Slaves
 - c. selbsttaktender Leitungscodierung → kann nur in eine Richtung senden, danach muss man die Master Position neu vergeben, um in die andere Richtung senden zu können → Latenzen treten durch Systemlaufzeiten auf → größte Fehleranfälligkeit (es muss bei jedem Schritt neu überlegt werden)

2. asynchrone Übertragung
 - a. in Echtzeit (z.B. Streaming, MP3-CD) → benötigt immer einen Puffer
 - b. nicht Echtzeit (z.B. CD Brennen)
- ob Echtzeit oder nicht entscheidet sich beim Produzent
- Dauer bleibt nicht erhalten → fällt meistens nicht auf
- bei Live muss der Kanal mehr durchlassen als das Signal an Bandbreite hätte

AES-3 Protokoll zur Audiosignalübertragung

- symmetrische Übertragung , XLR (Pin 1 Masse, Pin 2/3 Signal)
- Spannung 2-7V (Peak-to-Peak), minimal 200mV
- problemlos 100m und mehr an Kabellänge möglich

Aufteilung eines AES-3 Subframes:

- besteht aus 32 Bit:
 - o 0-3: Präambel: Kündigt das neue Subframe an
 - o 4-7: AUX - Bereich
 - o 8-11: Pad-Bits
 - o 12-27: 16 Bit Audiodaten
 - o 28: Validity Bit → zur Überprüfung des Signals
 - o 29: User Bit
 - o 30: Channel Bit → enthält Informationen wie Abtastrate, Wortbreite, Kumulation
 - o 31: Parity Bit → Paritätsbit
- um 24 Bit Audio zu Übertragen, werden Pad Bits und AUX Bereich mit genutzt
- LSB sitzt bei 16 Bit am Übergang zu den Pad Bits
- 16 Bit Receiver würde am LSB abschneiden
 - ➔ 16/24 Bit darf keinen Unterschied machen
- Codiert in BI-Phase Mark (außer Präambel)