

Quellencodierung

- Redundanz → z.B. Audiosignal in positiver Halbwelle → Vorzeichen kann man sich sparen
→ Reversible Redundanzreduktion → Eigenschaften der Quelle
- Man weiß schon vieles über das Signal, wenn man sich mit den Eigenschaften der Quelle aus einander setzt
→ Einzige mögliche Fehler sind Rundungsfehler
- Irrelevanzreduktion → Signalbereiche können „dünnere“ übertragen werden → zugeschnitten auf den Rezipienten (z.B. kein HD Fernsehen für mobilen Empfang)
- Es wird nur jenes Übertragen, was vom Rezipienten auch korrekt empfangen werden kann

Auftrittswahrscheinlichkeit eines Wortes

- Hinweis auf den Informationsgehalt eines Symbols (Bsp. Ja/Nein-Spiel)
- Das seltene Ereignis liefert viel Information, ein häufiges liefert kaum Information
- Informationsgehalt eines Symbols:

$$H(S_i) = \log_2 \left(\frac{1}{p_i} \right) = -\log_2(p_i)$$

p_i = Auftrittswahrscheinlichkeit eines Symbols

- Informationsgehalt aller Symbole → Summenbildung

$$H = \sum_{i=1}^N p_i \cdot H(S_i) = \sum_{i=1}^N p_i \cdot \log_2 \left(\frac{1}{p_i} \right) = -\sum_{i=1}^N p_i \cdot \log_2(p_i) \left[\frac{\text{bit}}{\text{Symbol}} \right]$$

→ Es liegt nicht immer 1 bit Information vor, sondern nur im seltensten Fall

- Bsp. binäres System → maximale Entropie bei $p = 0,5$
- Rechenbeispiel:
System mit 4 Zuständen

Symbol	p_i	Codierung mit fester Länge	Codierung mit optimierter Länge	$H(S_i)$
A	0,5	00	1	1
B	0,25	01	01	2
C	0,125	10	001	3
D	0,125	11	000	3

Berechnung des Informationsgehalts aller Symbole:

$$H = \sum_{i=1}^N p_i \cdot \log_2 \left(\frac{1}{p_i} \right) = 0,5 \cdot 1 + 0,25 \cdot 2 + 0,125 \cdot 3 + 0,125 \cdot 3 = 1,75$$

- Es werden bei fester Länge 2 Bit codiert, unabhängig von der Wahrscheinlichkeit, dass ein Symbol auftritt → jenes kann etwa 10% verbessert werden, wenn man als Grundlage für die Codierung die Häufigkeit der Symbole nutzt → Redundanzreduktion durch variable Längencodierung (z.B. mit Huffman-Code)

- Einzelne Bitfehler können dadurch größere Fehler verursachen:
 $\underline{1\ 01\ 001\ 000}$ → 4 Symbole erkannt
 $\underline{1\ 01}\quad \underline{01}\quad \underline{1\ 000}$ → 5 Symbole erkannt
die 1 ist ein Bitfehler
- Bsp. Akkord einer Jazzgitarre (analysiert in Matlab):
 - o Amplitudenbereich 0,8 – 1 kommt nicht vor
 - o Durch Funktion wird der Informationsgehalt berechnet → 12 bit (das Signal ist 16 Bit @ 44,1 kHz)
 - ➔ Nicht jedes Sample enthält 16 Bit Information
 - o Aus der Häufigkeitsverteilung geht hervor, dass 0 am häufigsten vorliegt, die Werte nah bei 0 auch recht häufig, höhere Werte sind sehr selten
 - o Differenzbildung (DSIG → Hochpassfilterung) → Aussteuerung verkleinert sich → Entropie liegt bei 9 bit
 - o Ein weiteres Differenzieren würde eine weitere Reduzierung auf 7 bit herbeiführen
 - ➔ Die Codierung bleibt verlustfrei, allerdings werden die Amplituden immer kleiner, deswegen wird die Fehlerwahrscheinlichkeit höher → Empfehlung: Höchstens 2-3 Mal differenzieren
- Bsp. Sägezahnspannung:
 - o Es kommt zwar jedes Symbol gleich oft vor, allerdings kann der Verlauf des Signals vorhergesagt werden, in den weiteren Perioden steckt keine neue Information
 - o Einzige Information ist die Grundschwingung, welche sich aus der Periodendauer ergibt
 - o Informationswert 1,1 (nur Information bei den Sprüngen)
 - o Ableitung besteht aus Gleichanteil & Diracstößen an den Pegelsprüngen → es steckt eigentlich nur ein Zeitimpuls als Information im Signal → extrem schlechte Entropie
- Bsp. Sinus:
 - o Information von 1,0 → es gibt eine „Spielregel“ → damit kann man für alle Ewigkeit fortsetzen
 - o Vorhersage muss für eine Redundanzreduktion möglich sein → Prädiktion dient immer als Vorgabe
 - o Erst wenn die „Spielregeln“ des Signalverlaufs bekannt sind, dann codiert werden

Kompression von Sprachsignalen (Vocoder):

- Nachbau der Sprache durch Pulsgenerator & Rauschgenerator + Filter mit variablen Koeffizienten
- Grundprinzip (s. differenzielle Codierung): Referenz bei der Erzeugung zur Optimierung, damit die Rekonstruktion besser funktioniert
- Sprache im Spektrogramm: Formanten (2-3 Harmonische erkennbar), Pitchshift
- Ermittlung von Filterkoeffizienten für kleine Zeitabstände (16-20 ms) → Kompression von 8-10 möglich
- System ist für Sprache entwickelt → Musik ist nicht gut zu erkennen
- Unterschied zwischen verschiedenen Sprachmustern hat Einfluss auf die Entwicklung des Systems (verschiedene Sprachen, Dialekte)
- Filter werden an Formantenkurven angepasst → Übergang zwischen den Zuständen des Filters muss beachtet werden → Treppenfunktionen sollten vermieden werden
- ➔ Frage der Erstellung des Tons / der Sprache, nicht des Hörens (Quelle!!)

- Bsp. Jones-Diagramm → Position der Zunge & Stärke des Luftstroms werden dargestellt → Vokale sind an den Rändern aufgetragen
→ Vokale können aus verschiedenen Sprachen erkannt werden
- Durch kategorisches Hören kann das Signal gestört sein, es wird aber trotzdem richtig verstanden
→ Forschung → Weg hin zur Zunge (Quelle) → Vocoder kennen die Zungenlage nicht

MPEG (Motion Pictures Expert Group) Audiocodierung:

- Man orientiert sich an den Bedürfnissen der Rezipienten
- MP3 → Maskierungsschwellen → Ruhehörpegel und stationäre Mithörschwelle
- Ansatz über Kanalkapazität → $C = \frac{SNR \cdot B}{3}$
- Gemisch von Tönen wird nicht komplett wahrgenommen → laute, dominante Töne verdecken leisere → Ermittlung von Mithörschwellen
- Weitere Möglichkeit der Reduktion sind sehr laute Passagen → danach nimmt das Ohr für 10 ms nichts wahr
- Quantisierungsrauschen wird bei dominanten Tönen auch nicht gehört → an SNR kann gespart werden
- Zerlegung in Frequenzbänder:
 - o 48 kHz bei 16 Bit
 - o Teilung in 32 Bänder mit 48 kHz bei 16 Bit (Faktor 32)
 - o Unterabtastung mit 1,5 kHz → 32 Bänder mit 1,5 kHz bei 16 Bit (Faktor 1/32)
- Unterabtastung → gleiche Entropie des Codes → Signalanteile werden durch Aliasing ins Basisband gespiegelt
- Durch Filterung bei der Decodierung wird ein im Basisband liegender, gespiegelter Anteil wieder auf seine ursprüngliche Frequenz zurückversetzt (aus TP kann man schnell HP entwickeln und umgekehrt)
- Maximalwertbestimmung: Legt den Amplitudenrahmen fest → Festlegung eines „Fensters“, in dem diese Maximalwerte gelten
- Psychoakustisches Modell entscheidet, wie viele Bit nötig sind, um das entsprechende Frequenzband zu codieren → jedes Band erhält eine eigene Wortbreite
- Beim Decodieren werden Bereiche mit 2 Bit mehr Quantisierungsrauschen enthalten → dieses wird allerdings im Normalfall durch einen Maskierer verdeckt und bleibt unhörbar
- Vergleich der 3 Layer:
 - o Layer 1: 512 Werte, 32 Bänder
 - o Layer 2: 1024 Werte, 32 Bänder
 - o Layer 3 (MP3): 576 Bänder, nicht-lineare Quantisierung, Huffman-Codierung, L/R – Redundanzreduktion
- Forschung im Bereich der menschlichen Konzentrationsfähigkeit → Was nimmt der Mensch wahr, wenn man mehr Quantisierungsrauschen hinzufügt? Merkt er den Unterschied?