

## Rechnernetze - Media Networking (WiSe 2023/2024)

### Übung 10

- 1) In einer fiktiven Video-Übertragung wird H.261 als Kodierungsverfahren verwendet, wobei jedes Paket (vereinfachend) genau ein Video-Frame enthält. Bei Paket 1 handelt es sich um ein I-Frame, alle weiteren Pakete sind (vereinfachend) P-Frames. Bei der Übertragung geht mitten im Strom ein Paket verloren. Daher soll bei der Anzeige einfach das vorige Frame erneut abgespielt werden.
- a) Welchen Effekt könnte das bei der Anzeige des Videos beim Empfänger haben, wenn es sich um ein Video-Telefonat handelt?
  - b) Könnte der Effekt anders aussehen, wenn es sich bei dem übertragenen Video um einen Spielfilm handeln würde?

Kurze Begründung.

(2 Punkte)

Bei der H.261-Kodierung werden die Videodaten mit einer Kombination aus I-Frames (Intra-Frames) und P-Frames (Predictive Frames) übertragen. I-Frames dienen als Synchronisationspunkte/Referenzbilder und stellen vollständige, unabhängige JPEG-kodierte Bilder dar, während P-Frames die Differenzen zu diesen Bildern enthalten. Dieser Ansatz ermöglicht eine effiziente Komprimierung, da nur die Änderungen zwischen aufeinanderfolgenden Bildern übertragen werden müssen.

Aus der Aufgabe geht hervor, dass ein Paket in der Mitte des Datenstroms verloren gegangen ist und dass jedes Paket genau ein Videobild überträgt. Da nur das erste Paket ein I-Frame ist und alle anderen Pakete P-Frames sind, kann gefolgert werden, dass es sich bei dem verlorenen Paket um ein P-Frame handelt.

Im Kontext eines Videotelefonats könnte der Verlust eines P-Frames in der Regel kaum wahrnehmbar sein. Dies liegt zum einen daran, dass die Anzahl der Paketverluste zu gering ist, um einen spürbaren Qualitätsverlust zu verursachen. Zweitens finden während eines Videotelefonats normalerweise nur begrenzte Bewegungen statt, z.B. Lippen- und Augenbewegungen, die für den Gesprächspartner weitgehend irrelevant sind. Daher können die geringe Anzahl von Bildänderungen und die kontinuierliche Wiedergabe des vorherigen Frames den Effekt für den Empfänger minimieren.

Im Gegensatz dazu könnte der Verlust bei einem Spielfilm auffällig sein. Dies resultiert daraus, dass Spielfilme oft zahlreiche Szenenwechsel aufweisen, welche intensive Bildänderungen darstellen und somit als I-Frames kodiert werden sollten. In dem vorliegenden Fall wurde nur ein einziger I-Frame übertragen, der Rest als P-Frames. Wenn man nun davon ausgeht, dass ein P-Frame, der einen Szenenwechsel repräsentiert,

Eher untertrieben: Die Folge-Frames würden dann ja ggf. zu völlig anderen Bildern dekodiert werden.

verloren geht, kann dies zu einem kurzzeitigen Ruck oder einem sichtbaren Sprung in der Bewegung führen.

1,5 Punkte

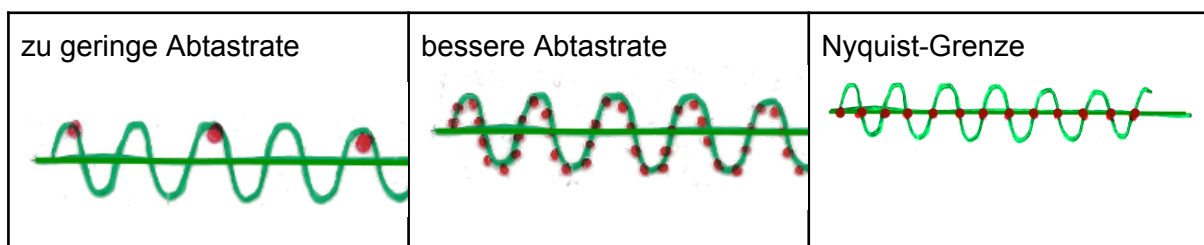
- 2) Eine fiktive Audio-Anwendung soll einen Frequenzbereich von 100 Hz bis 4,8 kHz unterstützen. Für die RTP-basierte Übertragung wird der folgende Protokollstack verwendet. Das verwendete fiktive Funknetz arbeitet mit Paketen von 100 Bytes Gesamtlänge, davon 10 Bytes Header. Die RTP-Pakete haben einen minimalen Header (von 12 Bytes). IP-Optionen, IP-Fragmentierung und *Header Compression* werden nicht verwendet. Für den Paket-Strom dieser Anwendung stehen insgesamt 80 kbit/s zur Verfügung.

RTP
UDP
IPv4
Subnetz

- Welche Abtastrate sollte das fiktive Audio-Kodierungsverfahren verwenden?
  - Das fiktive Audio-Kodierungsverfahren verwendet ein DPCM-Verfahren. Wieviele Bits stehen für die einzelnen Abtastwerte zur Verfügung?
  - Wie gut funktioniert dieses Verfahren bei Paketverlusten?
- Jeweils mit kurzer Begründung.

(3 Punkte)

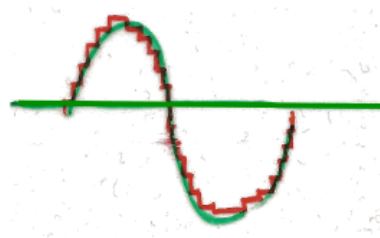
Bei Audiosignalen handelt es sich um Schwingungen, die durch wechselnde Luftdichte entstehen, also z.B. erzeugt durch eine schwingende Lautsprechermembran. Dabei bestimmt die Frequenz der Schwingung, die Tonhöhe und die Amplitude die Lautstärke. In der Regel hat man es mit übereinander gelagerten Schwingungen zu tun. Um nun ein Audiosignal digital kodieren zu können, muss man den aktuellen Wert regelmäßig abgreifen. Dabei ist eine Frage wesentlich: Wie hoch muss die Abtastrate sein? Unten aufgeführt sind verschiedene hohe Frequenzen von Abtastraten, wodurch Schwingungsverläufe besser oder schlechter vermutet werden können.



Nun kann es aber sein, dass zwischen je zwei Abtastwerte eine oder mehr weiterer, komplette Schwingungen höherer Frequenz liegen. An dieser Stelle greift die Nyquist-Grenze (ganz rechts, oben in der Tabelle): Die Abtastrate muss **mehr** als doppelt so

groß sein wie die höchste Frequenz, die man dadurch darstellen können will. Das Beispiel ganz rechts macht außerdem noch deutlich, warum 2 Abtastwerte noch nicht reichen, um den eigentlichen Verlauf zu rekonstruieren.  $> 2$  Abtastwerte pro Schwingung reichen i.d.R. aus. Allerdings gilt grundsätzlich, dass je mehr Abtastwerte es gibt, dass auch die Schwingung besser nachgebildet wird.

Abtastwerte: Die Anzahl der Abtastwerte werden in digitalen Audio-Kodierungsverfahren digital kodiert (siehe unten).



nein, mehr als doppelt so hoch

Mit  $n$  Bits  $\rightarrow$  kann man  $2^n$  verschiedene Amplituden unterscheiden. Es findet eine Quantisierung statt. Bei entsprechend vielen Abtastwerten pro Schwingung könnte das so ähnlich wie im Bild oben aussehen.

Zusammenfassend lässt sich sagen, dass die Abtastrate (Anzahl der Samples pro Sekunde) von dem zu überdeckenden Frequenzspektrum abhängt. Hierbei sollte die Abtastrate nach dem Nyquist-Theorem mindestens doppelt so hoch sein wie die höchste zu übertragende Frequenz, um eine korrekte Wiedergabe zu gewährleisten. Des Weiteren ist die Anzahl der Bits pro Sample (Auflösung) abhängig von der gewünschten Genauigkeit der Audioübertragung. Eine höhere Auflösung, also mehr Bits pro Sample, ermöglicht eine präzisere Darstellung des Audiosignals, jedoch geht dies mit einer höheren Datenrate einher.

a) Das Nyquist-Shannon-Theorem besagt, dass die Abtastfrequenz eines Signals mindestens doppelt so hoch sein muss wie die höchste Frequenzkomponente des Signals, um eine fehlerfreie Rekonstruktion des Originalsignals aus den abgetasteten Daten zu ermöglichen. Anders ausgedrückt: Damit keine Informationen verloren gehen, muss die Abtastfrequenz doppelt so hoch sein wie die höchste Frequenz des Signals.

In dem vorliegenden Fall ist die höchste Frequenz des Signals 4,8 kHz. Die minimale Abtastfrequenz (für ein ungefähres Minimum) wäre also  $2 \times 4,8 \text{ kHz} = 9,6 \text{ kHz}$ . Die Abtastfrequenz sollte also mindestens 9,6 kHz betragen.

b) DPCM nach G.711, 8 kHz Abtastrate: Hierbei werden die Abtastwerte nicht absolut kodiert, sondern als Differenz zum vorherigen Wert. Da sehr große Sprünge von einem Wert zum anderen sehr unwahrscheinlich sind, werden dabei kleine Werte verwendet, die sehr kompakt kodiert werden können. Weiterhin wird eine weitere Komprimierung durch ADPCM erreicht. Hierbei wird berücksichtigt, dass sich das ganze Schwingungsmuster während eines Lauts kaum ändert. Statt einzelne Abtastwerte immer wieder zu kodieren, kann auch ein Code dafür verwendet werden, dass die nächste Schwingung so ist wie die vorige (G.721, 8 kHz Abtastrate, 32 kbit/s). G.711/G.721 sind zwar für Telefonverkehr geeignet, aber der überdeckte Frequenzbereich ist zu klein - für einerseits gute Präsenz (Nebengeräusche o.Ä.) und höhere Frequenzen für z.B. Sprechererkennung (bei Gruppenkommunikation etwas schwierig). Dafür hat man einen weiteren Codec entwickelt: G.722:  $\approx 50 \text{ Hz} - 7, \dots \text{ kHz}$  (+ ADPCM). Da die niedrigen Frequenzen für die Sprachverständlichkeit besonders wichtig sind, werden sie genauer kodiert als die höheren Frequenzen. Dieser Standard kann mit verschiedenen Datenraten arbeiten:

$\geq 4 \text{ kHz}$ : 2 Bit/Sample  
 $< 4 \text{ kHz}$ : 6 Bit/Sample  $\rightarrow 64 \text{ kbit/s}$   
5 Bit/Sample  $\rightarrow 56 \text{ kbit/s}$   
4 Bit/Sample  $\rightarrow 48 \text{ kbit/s}$  ( $\triangleq$  G.721)

Insgesamt kommt man hier auf Datenraten von 48 kbit/s bis 64 kbit/s.

Für die Berechnung der Bits pro Sample verwenden wir eine Quantisierung von 16. Die Quantisierung bezieht sich auf die Anzahl der diskreten Werte, die ein analoges Signal repräsentieren können. Wenn wir von einer Quantisierung von 16 sprechen, bedeutet das, dass das kontinuierliche analoge Signal in 16 verschiedene Intervalle oder Schritte unterteilt wird. Diese 16 Intervalle repräsentieren die diskreten Werte, die für jeden Abtastwert ausgewählt werden können.

In einem digitalen Signalverarbeitungssystem, wie es bei DPCM der Fall ist, werden die kontinuierlichen Abtastwerte des analogen Signals in diskrete digitale Werte umgewandelt. Die Anzahl der möglichen digitalen Werte wird durch die Quantisierung bestimmt. In diesem

Kontext würde eine Quantisierung von 16 bedeuten, dass jeder Abtastwert auf einen von 16 möglichen digitalen Werten abgebildet wird.

Die Formel zur Berechnung der Anzahl der Bits pro Abtastwert lautet:

$$\text{Bits pro Abtastwert} = \text{Bittiefe} = \log_2(\text{Quantisierung})$$

In diesem Fall wäre die Bittiefe  $= \log_2(16) = 4$ , was bedeutet, dass 4 Bits benötigt werden, um jeden Abtastwert zu repräsentieren.

Die Wahl der Quantisierung beeinflusst die Genauigkeit der Signalrepräsentation. Eine höhere Quantisierung ermöglicht eine genauere Abbildung des analogen Signals in den digitalen Raum, benötigt jedoch auch mehr Bits pro Abtastwert und kann daher mehr Bandbreite erfordern.

Wir haben nun eine Abtastrate von 9,6 kHz und wollen die für Nutzdaten verfügbare Bandbreite von 40 kbit/s berücksichtigen. Die Bandbreite für Nutzdaten allein ist wie folgt berechnet worden:

$$\text{Nutzdaten pro Paket} = 100 \text{ Bytes (Gesamtlänge)} - 10 \text{ Bytes (Header des Pakets)} - 12 \text{ Bytes (RTP-Header)} - 20 \text{ Bytes (min. IPv4-Header)} - 8 \text{ Bytes (UDP-Header)} = 50 \text{ Bytes}$$

Insgesamt steht uns eine Bandbreite von 80 kbit/s = 10 000 Bytes/s zur Verfügung. Daraus lässt sich errechnen, wie viele Nutzdaten pro Sekunde übertragen werden können:  $10\,000 \text{ Bytes/s} / 100 \text{ Bytes} = 100 \text{ Pakete}$  mit je 50 Bytes Nutzdaten, d.h.  $100 \times 50 \text{ Bytes} = 5\,000 \text{ Bytes/s} = 40 \text{ kbit/s}$ .

Nun können wir sicherstellen, dass die Anzahl der Bits, die pro Sekunde für die Darstellung eines Audiosignals verwendet werden, die berechnete Bandbreite nicht überschreitet:

$$\text{Bitrate} = \text{Frequenz} \times \text{Bittiefe} \times \text{Kanäle}$$

$$\text{Bitrate} = 9,6 \text{ kHz} \times 4 \text{ Bits} \times 1 = 38,4 \text{ kbit/s}$$

Es wird also eine Bandbreite von 38,4 kbit/s benötigt, was unter der verfügbaren Bandbreite von 40 kbit/s liegt.

c) Wie bereits bei b) thematisiert, geht es hier nochmal um DPCM-Verfahren

**DPCM:** Bei DPCM kann ein einzelner Paketverlust zu einer kumulativen Fehlerakkumulation führen. Da jeder Abtastwert auf dem vorherigen basiert, kann der Verlust eines Pakets die korrekte Rekonstruktion der nachfolgenden Abtastwerte beeinträchtigen. Ohne geeignete Maßnahmen zur Fehlerkorrektur oder zum Schutz vor Paketverlusten kann DPCM anfällig für Qualitätseinbußen sein. ADPCM kann robuster gegenüber Paketverlusten sein als DPCM (Differenzierung oben (b)) bereits geschehen), da es eine adaptive Anpassung der Quantisierungsschrittgröße bietet. Die Anpassung ermöglicht es, sich besser an unterschiedliche Signalcharakteristiken anzupassen, was die Auswirkungen von Paketverlusten auf aufeinanderfolgende Abtastwerte minimieren kann. Erwähnenswert ist auch, dass DPCM gegenüber PCM bei vergleichbaren Signalstörungen bessere Rekonstruktionsergebnisse erzielen kann, wenn auch nur einige der eingesparten Bits für die Fehlerkorrektur verwendet werden. Die Fehlerkorrektur muss jedoch immer auf alle Abtastwerte des DPCM-Signals in gleicher Weise angewendet werden<sup>1</sup>.

RTP-basierte Übertragung und Paketverluste: eine der Aufgaben von RTP ist es, eine sinnvolle Aufteilung der Informationsströme auf Pakete. Dabei wird auch der Quereinstieg/Wiederaufsetzen nach Paketverlusten berücksichtigt.

Die Unterschiede zwischen Audio- und Videostreams zeigen sich bei den verwendeten Verfahren. Audio Samples sind meistens relativ kurz, daher ist es naheliegend, eine ganzzahlige Anzahl solcher Samples pro Paket zu transportieren. Ein Problem bei der Übertragung sind die oft variierenden Übertragungszeiten in paketvermittelten Netzen ohne Reservierungsverfahren (Jitter).

RTP Pakete enthalten weiterhin eine stromspezifische Sequenznummer. Auf diese Weise lassen sich Paketverluste erkennen und auch falsche Empfangsreihenfolgen eines Paketstroms. Da es hier um die Übertragung von zeitabhängige Informationsströmen geht, spielt insbesondere die Funktionalität der Zeitsynchronisation in RTP eine Rolle, dabei

---

<sup>1</sup> [https://link.springer.com/content/pdf/10.1007/978-3-642-79359-2\\_12.pdf](https://link.springer.com/content/pdf/10.1007/978-3-642-79359-2_12.pdf)

können drei Formen unterschieden werden: Abspielsynchronisation - müssen in genau der richtigen Taktrate abgespielt werden, wie sie auch beim Sendesystem aufgezeichnet worden waren. Die zweite Synchronisationsform, Intermediasynchronisation, die zeitliche Zusammenhänge zwischen verschiedenen Informationsströmen ausdrückt, wie z.B. Synchronisation zwischen Audio und Videospur, die bei RTP durch zwei verschiedene Paketströme realisiert werden. Die dritte Form der Synchronisation kommt ins Spiel, wenn es Informationsströme verschiedener Teilnehmer innerhalb einer Konferenz geht, wobei es wichtig sein kann, wann die Teilnehmer jeweils ihre Nachrichten gesendet haben, welche globale Reihenfolge sie also haben. Die Übertragungsvorgänge nehmen eine variable Zeit in Anspruch; der RTP-Anteil an dieser Lösung ist einfach, wobei in den Paketen ein Zeitstempel mitgeschickt wird (Sendezeit), die beim Empfängersystem entsprechend interpretiert werden kann. Für die ersten beiden hier genannten Synchronisationsformen reichen diese Zeitstempel aus, um die empfangenen Pakete zeitlich richtig zu synchronisieren. Im Zusammenhang mit der dritten Synchronisationsform benötigt man zusätzlich z.B. NTP (synchronisierte Uhren in den beteiligten Sendesystemen). Einen absoluten Zeitstempel aber erhält nicht mal das (erste) RTP selber, sondern die "reale" Zeit beim Sender wird durch ein begleitendes RTCP-Steuerpaket abgebildet. Alle weiteren Pakete des RTP Stroms enthalten Zeitdifferenzen zum Vorgängerpaket. Das erlaubt eine kompaktere Kodierung - hat aber den Nachteil, dass die dabei Einheiten mit der realen Samplingrate des erzeugenden Geräts zusammenpassen müssen. Relative Zeitangaben haben außerdem das Problem, dass sie Timingfehler bei Paketverlusten erzeugen können (sind aber durch die Sequenznummern identifizierbar + gleichmäßige Samplingrate).

2,5 Punkte