

Tentamen i  
**Kompression av ljud och bild**  
**TSBK35**

24:e mars 2022 14:00 - 18:00

<b>Lokal:</b>	SP71, SP34, SP303
<b>Examinator:</b>	Harald Nautsch
<b>Lärare:</b>	Harald Nautsch, 0701718715
<b>Institution:</b>	ISY
<b>Provkod:</b>	TEN1
<b>Antal uppgifter:</b>	9
<b>Antal sidor:</b>	4 + formelsamling
<b>Hjälpmedel:</b>	Miniräknare, "Tables and Formulas for Image Coding and Data Compression"
<b>Betyg:</b>	0-13 U 14-19 3 20-25 4 26-30 5
<b>Övrigt:</b>	Läraren besöker tentalokalen ca 15.15 och 16.45

- 1 Vid kodning av video och stillbilder i färg brukar man använda färgrymden YCbCr istället för RGB. Förklara hur färgrymderna skiljer sig åt och varför man föredrar YCbCr.

(2 p)
  
- 2 Förklara (ganska detaljerat) hur en modern *hybridkodare* för video fungerar. Beskriv både kodning och avkodning.

(4 p)
  
- 3 a) Beskriv några fördelar och nackdelar med att använda *vektorkvantisering* istället för *skalär kvantisering*.

(2 p)

  
b) Förklara vad *LBG-algoritmen* är och hur den fungerar.

(2 p)
  
- 4 Det mänskliga hörselsinnet har vissa egenskaper som kan utnyttjas för att få en effektivare kodning av ljudsignaler. Förklara vilka dessa egenskaper är och hur de kan utnyttjas i kodningen.

(4 p)
  
- 5 I vanlig metod för talkodning är CELP. Ge en kortfattad beskrivning av hur den fungerar, både på kodar- och avkodarsidan.

(2 p)

6 En minnesfri källa har alfabetet  $\mathcal{A} = \{x, y, z\}$  och symbol-sannolikheterna  $P(x) = 0.5$ ,  $P(y) = 0.4$  och  $P(z) = 0.1$ .

a) Vilken är den teoretiskt lägsta medeldataakten man kan få vid distorsionsfri kodning av källan?

(1 p)

b) Konstruera en huffmankod för par av symboler från källan och beräkna dess medeldataakt (i bitar/symbol).

(2 p)

7 En källa har alfabetet  $\{a, b, c, d, e, f, g, h\}$ . En lång sekvens från källan kodas med LZW. Den resulterande indexsekvensen börjar som

6, 0, 8, 10, 0, 1, 4, 14, 14, 3, 9, 12, ...

Startordboken är:

index	sekvens	index	sekvens
0	<i>a</i>	4	<i>e</i>
1	<i>b</i>	5	<i>f</i>
2	<i>c</i>	6	<i>g</i>
3	<i>d</i>	7	<i>h</i>

Avkoda indexsekvensen så långt som möjligt. Ange även hur den slutliga ordboken ser ut.

(3 p)

- 8 En bildkälla modelleras som en tvådimensionell normalfördelad process  $X_{i,j}$  (där  $i$  och  $j$  är koordinater i bilden) med medelvärde noll. Autokorrelationsfunktionen

$$R_{XX}(k, l) = E\{X_{i,j} \cdot X_{i+k,j+l}\}$$

har från en stor mängd testdata skattats till:

$$\begin{aligned} R_{XX}(0, 0) &= 251, & R_{XX}(1, 0) &= 227 \\ R_{XX}(0, 1) &= 222, & R_{XX}(1, 1) &= R_{XX}(1, -1) = 205 \end{aligned}$$

Källan kodas med prediktorn

$$p_{i,j} = a_1 \cdot \hat{X}_{i-1,j} + a_2 \cdot \hat{X}_{i,j-1}$$

Hitta de prediktorkoefficienter  $a_1$  och  $a_2$  som minimerar prediktionsfelets varians.

Antag att prediktionsfelet kvantiseras likformigt och kodas med en aritmetisk kodare. Steglängden i kvantiseraren väljs så att datatakten efter källkodning blir 4.5 bitar/sampel. Vad blir det resulterande signal-brusförhållandet i dB?

(4 p)

- 9 En ljudsignal modelleras som en tidsdiskret stationär normalfördelad process  $X_n$  med medelvärde 0 och autokorrelationsfunktion  $R_{XX}(k) = E\{X_n \cdot X_{n+k}\} = 0.94^{|k|}$ .

Signalen transformkodas med en 4-punkters DCT. Transformkomponenterna Lloyd-Max-kvantiseras så att medeldataakten blir 1.75 bitar/sampel.

Fördela bitar till transformkomponenterna så att medeldistorsionen minimeras och beräkna det resulterande signal-brus-förhållandet.

(4 p)